frontiers | Research Topics

Polyploidy and its consequences

Edited by
Yves Van de Peer, Andrew H. Paterson and Jonathan F. Wendel

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-3445-8
DOI 10.3389/978-2-8325-3445-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Polyploidy and its consequences

Topic editors

Yves Van de Peer — Ghent University, Belgium
Andrew H. Paterson — University of Georgia, United States
Jonathan F. Wendel — Iowa State University, United States

Citation

Van de Peer, Y., Paterson, A. H., Wendel, J. F., eds. (2023). *Polyploidy and its consequences*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3445-8

Table of contents

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 May 2020
doi: 10.3389/fgene.2020.00317

The Evolution of an Invasive Plant, *Sorghum halepense* L. ('Johnsongrass')

Andrew H. Paterson[1], WenQian Kong[1], Robyn M. Johnston[2], Pheonah Nabukalu[3], Guohong Wu[4], William L. Poehlman[5], Valorie H. Goff[1], Krista Isaacs[6], Tae-Ho Lee[1,7], Hui Guo[1], Dong Zhang[1], Uzay U. Sezen[1], Megan Kennedy[4], Diane Bauer[4], Frank A. Feltus[5], Eva Weltzien[6,8], Henry Frederick Rattunde[6,8], Jacob N. Barney[9], Kerrie Barry[4], T. Stan Cox[3] and Michael J. Scanlon[2]*

[1] Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, United States, [2] School of Integrative Plant Science, Cornell University, Ithaca, NY, United States, [3] The Land Institute, Salina, KS, United States, [4] Department of Energy Joint Genome Institute, Walnut Creek, CA, United States, [5] Department of Genetics & Biochemistry, Clemson University, Clemson, SC, United States, [6] International Crops Research Institute for the Semi-Arid Tropics, Bamako, Mali, [7] Genomics Division, National Institute of Agricultural Sciences, Jeonju, South Korea, [8] College of Agricultural and Life Science, University of Wisconsin-Madison, Madison, WI, United States, [9] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States

OPEN ACCESS

Edited by:
Vijay Kumar Tiwari, University of Maryland, College Park, United States

Reviewed by:
Ze Peng, University of Florida, United States
Amita Mohan, University of Pennsylvania, United States

***Correspondence:**
Andrew H. Paterson
paterson@uga.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *08 November 2019*
Accepted: *17 March 2020*
Published: *14 May 2020*

Citation:
Paterson AH, Kong W, Johnston RM, Nabukalu P, Wu G, Poehlman WL, Goff VH, Isaacs K, Lee T-H, Guo H, Zhang D, Sezen UU, Kennedy M, Bauer D, Feltus FA, Weltzien E, Rattunde HF, Barney JN, Barry K, Cox TS and Scanlon MJ (2020) The Evolution of an Invasive Plant, Sorghum halepense L. ('Johnsongrass'). Front. Genet. 11:317. doi: 10.3389/fgene.2020.00317

From noble beginnings as a prospective forage, polyploid *Sorghum halepense* ('Johnsongrass') is both an invasive species and one of the world's worst agricultural weeds. Formed by *S. bicolor* x *S. propinquum* hybridization, we show *S. halepense* to have *S. bicolor*-enriched allele composition and striking mutations in 5,957 genes that differentiate it from representatives of its progenitor species and an outgroup. The spread of *S. halepense* may have been facilitated by introgression from closely-related cultivated sorghum near genetic loci affecting rhizome development, seed size, and levels of lutein, a photochemical protectant and abscisic acid precursor. Rhizomes, subterranean stems that store carbohydrates and spawn clonal propagules, have growth correlated with reproductive rather than other vegetative tissues, and increase survival of both temperate cold seasons and tropical dry seasons. Rhizomes of *S. halepense* are more extensive than those of its rhizomatous progenitor *S. propinquum*, with gene expression including many alleles from its non-rhizomatous *S. bicolor* progenitor. The first surviving polyploid in its lineage in ~96 million years, its post-Columbian spread across six continents carried rich genetic diversity that in the United States has facilitated transition from agricultural to non-agricultural niches. Projected to spread another 200–600 km northward in the coming century, despite its drawbacks *S. halepense* may offer novel alleles and traits of value to improvement of sorghum.

Keywords: invasion biology, polyploidy, evolutionary novelty, weed, crop, rhizome, perennial

INTRODUCTION

Cytological, morphological (Celarier, 1958; Doggett, 1976), and molecular data (Paterson et al., 1995) suggest that tetraploid *Sorghum halepense* (2n = 40) arose as a naturally occurring hybrid between *S. bicolor* (2n = 20), an annual, polytypic African species which includes cultivated sorghum; and *S. propinquum* (2n = 20), a perennial southeast Asian native of moist habitats. While

a firm estimate of its antiquity is lacking, *S. propinquum* is thought to have shared ancestry with *S. bicolor* ∼1–2 million years ago (Feltus et al., 2004), roughly circumscribing the maximum age of *S. halepense*. Occasionally used as forage and even food (seed/flour), *S. halepense* has spread in post-Columbian times from its hypothesized west Asian center of origin across much of Asia, Africa, Europe, North and South America, and Australia. Its establishment in the U.S. is probably typical of its spread to other continents, being introduced intentionally as a prospective forage and unintentionally as a contaminant of seedlots (McWhorter, 1971). However, while sorghum largely remained confined to cultivation, *S. halepense* readily naturalized and has spread across much of North America, both to agricultural and non-agricultural habitats (Sezen et al., 2016) – suggesting capabilities for adaptation well beyond those of sorghum.

Its common name thought to be a misnomer [the eponymous Col. Johnson may have obtained propagules from his wife's family, who accidentally introduced it to South Carolina shortly after the Revolutionary War (Tellman, 1996)], 'Johnsongrass' has the rare distinction of being both a noxious weed in 20 U.S. states and an invasive species in 16 (Quinn et al., 2013). With at least 24 herbicide-resistant biotypes now known (Heap, 2012), Johnsongrass appears likely to become even more problematic in the future. For example, a glyphosate resistant biotype discovered in Argentina in 2002 covered 10,000 ha by 2009 (Binimelis et al., 2009). Its ability to cross with sorghum despite a ploidy barrier (reviewed in Warwick and Black, 1983; Tang and Liang, 1988) makes Johnsongrass a paradigm for the dangers of crop 'gene escape,' and restricts deployment of many transgenes that could reduce the cost and increase the stability of sorghum production.

Here, we integrate several diverse data types to elucidate the evolution of *S. halepense*, its invasiveness as exemplified by rapid spread across the United States in post-Columbian times, and the roles of polyploidy and interspecific hybridity in distinctive features of its growth and development. As the first surviving polyploid in its lineage in ∼96 million years (Paterson et al., 2009; Wang et al., 2015), *S. halepense* may also open new doors to sorghum improvement, with synergy between gene duplication and interspecific hybridity nurturing the evolution of genes with new or modified functions (Ohno, 1970).

MATERIALS AND METHODS

Genome Size Determination

Sorghum halepense genome size is an average for five accessions based on flow cytometry performed on a fee for service basis under the supervision of K. Arumuganathan, Benaroya Research Institute, using published methods (Arumuganathan and Earle, 1991).

Resequencing

Sorghum halepense, *S. propinquum*, *S. timorense* and representatives of each of the wild *S. bicolor* races (*S. bicolor* ssp. drummondii, SRP116974; *S. bicolor* ssp. verticilliflorum race aethiopicum, SRP116975: *S. bicolor* ssp. verticilliflorum race arundinaceum, SRP116973; *S. bicolor* ssp. verticilliflorum race verticilliflorum, SRP116978: *S. bicolor* ssp. verticilliflorum race virgatum SRP116940) were sequenced using standard methods implemented at the US Department of Energy Joint Genome Institute, as part of a larger project including 27 genomes and 39 transcriptomes in total. From each accession, 76-bp paired-end reads were aligned to the *Sorghum bicolor* reference genome (v1.4) using BWA version 0.5.9 (Li and Durbin, 2009). Multiple-sample SNP calling was performed using the mpileup program in the samtools package and bcftools (Li et al., 2009). Reads with mapping quality score > = 25 and base quality > = 20 are used for SNP calling. Raw SNPs are further filtered according to read depth distribution to avoid paralog contamination and low coverage regions. Each accession's genotype is calculated using maximum likelihood estimation using reads with coverage between 4 and 30X. The genotype with the largest likelihood is assigned to each individual. SNPs with allele frequency > = 0.01 are used for downstream analysis.

As tandem genes are often recently derived and share high sequence similarity, they can complicate short read alignment and introduce 'false SNPs' from paralogs. To address this, the coverage of genomic reads (not including transcriptome data) was examined for every tandem gene in the sorghum genome. The average coverage of the whole genome across the 27 genomes studied is about 553X. There were 31 tandem genes with more than twice the genome coverage (1100X), of which 7 have coverage more than 2500X (ranging up to 7500X). A total of 14 of the 31 high coverage tandem genes have SNPs called, and were removed from further analysis.

SNP Inference

To identify *S. halepense* SNPs, reads from *S. halepense* were aligned to the reference *S. bicolor* genome by BWA and SNPs determined with nucleotide groups for each reference *S. bicolor* genomic position by an in-house script. False positive *S. halepense* SNPs for each position of the reference *S. bicolor* genome were inferred and removed, based on three criteria: (i) if the top two nucleotide groups are the same as reference *S. bicolor* and *S. propinquum*, respectively, there are no false positive SNPs; (ii) if read depth of an SNP is 1 (noting the average 14X coverage of the *S. halepense* genome), a false positive was inferred; (iii) if *p*-value calculated by the Fisher exact test for the actual and theoretical read depths (bicolor:propinquum is 1:1), is less than 0.1, a false positive was inferred. The full SNP table with the reference *S. bicolor*, *S. propinquum*, and *S. halepense* SNPs as well as wild *S. bicolor* and *S. timorense* SNPs determined with total RNA and genomic DNA, respectively, against the reference *S. bicolor* genome, is provided (**Supplementary Table 1**). Classifications of duplicated genes into paralogs versus homologs followed the *S. bicolor* reference genome (Paterson et al., 2009).

Gene Functional Enrichment Analysis

Arabidopsis GO-slim gene annotation was used for function enrichment analysis. GO-slim terms are assigned to sorghum genes based on sequence similarity inferred from best blastp hit. Binomial distribution based on the proportion of a GO-slim term among all annotated genes in the sorghum genome is used as

the null distribution. Test significance threshold is defined as $p < 0.05$, unless specified otherwise.

Functional Impact of SNPs

A customized script is used to map SNPs to the *Sorghum bicolor* gene model version 1.4. Striking SNPs are identified as those mapped to coding regions, splicing sites, stop codons and transcription initiation sites. The functional impact of non-synonymous SNP is assessed based on the evolutionary conservation profile of amino acids. Orthologous groups of protein sequences from 30 plant species are constructed using OrthoMCL. Protein sequences from each orthologous group are aligned using Clustalw2 (Larkin et al., 2007). Non-synonymous SNPs are mapped to the alignment of the corresponding orthologous group and a 'functional impact score' is calculated with a modified entropy function (Reva et al., 2011):

$$S_i(\alpha \to \beta) = \begin{cases} -\ln \frac{n_i(\beta)+1}{n_i(\alpha)} P_c & \text{if } n_i(\beta)+1 < n_i(\alpha) \\ -\ln \frac{n_i(\beta)+1}{n_i(\alpha)} (1-P_c) & \text{if } n_i(\beta)+1 > n_i(\alpha) \end{cases}$$

where, α, β are 20 amino acid residues and gaps, $n_i(\alpha)$ is the number of occurrences of residue α in an alignment column *i*. $n_i(\beta)$ is the number of occurrences of an alternative residue β in the column *i*. *Pc* is the probability of occurrence of the most common residue in the alignment column *i*. Si is the function index score, a measure of functional impact of a mutation on protein function. The significance threshold of S_i is determined at the FDR = 0.01.

Survival of Cold or Dry Seasons

Survival of cold (temperate) or dry (Mali) seasons was based on single plants (*SbxSh* F2), or at least some survival within progeny rows of about 5 (*SbxSp* RILs) or 10 plants (*SbxSh* BC1F2; F3). Methods for determining rhizome numbers and distances from the originating crown are as cited (Kong et al., 2015). Flowering time was based on the average number of days from planting to flowering of either single plants (*SbxSh* F2) or the first five plants in a plot, and vegetative biomass was determined at the end of the growing season (after frost) by harvesting all tissue >2 cm above the ground for entire plots, separating inflorescences from vegetative tissues, drying to stable mass, and determining dry tissue mass. Heritabilities were calculated from F_2–F_3 regression (*S. bicolor* x *S. halepense* F_2), or variance component analysis [*S. bicolor* x *S. halepense* BC_1F_2; *S. bicolor* x *S. propinquum* RILs (Kong et al., 2015)].

A logistic regression was performed using dry-season survival by each genotype as the response variable (0 or 1) and the distance between rhizome derived shoots and the crown based data from Athens, GA in 2013 ('Dist'), as the explanatory variable. The model is:

$$ln(p/(1-p)) = -2.1358 + 0.2339 \times Dist,$$

where p is the probability of survival. With 1 cm increments in Dist, the probability of survival increased by 3%.

Laser Microdissection RNA-seq (LM RNA-seq)

LM RNA-seq was used to compare transcript accumulation in the shoot apices of buds induced to develop as either secondary rhizomes or leafy shoots (**Figure 1**). The meristem plus two youngest leaf primordia were microdissected from transverse sections. Two replicates of each meristem type were collected, with 5 meristems per replicate. LM, RNA extraction and amplification, cDNA library preparation and Illumina sequencing were performed as described (Takacs et al., 2012). LM RNA-seq reads are archived under NCBI BioProject ID PRJNA356741.

Specificity of Gene Expression

Sorghum halepense RNAseq FASTQ files were preprocessed with (Bolger et al., 2014) (0.22) and assembled into a transcriptome reference assembly using Trinity (Haas et al., 2013) (r06-08-2012; –kmer_method jellyfish). Transcript mapping to the reference sorghum genome, and differential gene expression was performed with TopHat (Trapnell et al., 2009) (v2.0.3), Bowtie2 (Langmead and Salzberg, 2012) (2.0.0.6), Samtools (Li et al., 2009) (0.1.18.0), and Cufflinks (Trapnell et al., 2012) (2.0.0). From the FPKM values in **Supplementary Table 4**, three gene lists were created: (1) Significant differentially expressed genes between shoot and rhizome; (2) Genes ON in rhizome and OFF in Shoot; (3) Genes ON in shoot and OFF in rhizome. The rank order of differentially expressed genes was based on the cuffdiff test statistic (Trapnell et al., 2010), which was very closely correlated with the fold change in gene expression (**Supplementary Table 5**). To annotate these lists, the most recent *S. bicolor* reference genome annotation (v3.1) was downloaded from Phytozome v11.0 and annotation

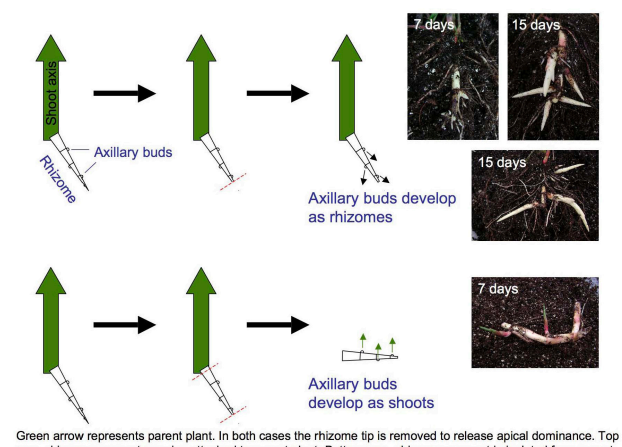


FIGURE 1 | Excising primary rhizome tips induces *Sorghum halepense* axillary bud growth. Buds on rhizomes attached to the parent shoot (large green arrows) develop as secondary rhizomes (above), whereas buds on excised rhizomes develop as leafy shoots (below). Mechanical excision ensured the identity and equivalent developmental staging of shoot apices selected for transcriptional profiling.

labels for GO, KEGG, and PFAM and were assigned to the *S. halepense* transcripts via homolog mapping with BLASTN. Genes were categorized as ON if there was any expression detected (FPKM > 0), and OFF if the FPKM value was zero. Term enrichment was performed using the David (da Huang et al., 2007) method re-implemented in a Perl script where the gene background was limited to a non-redundant list of *S. bicolor* transcripts that mapped to the Trinity transcript IDs from **Supplementary Table 4**.

Correspondence of Sorghum QTLs to Introgression Hotspots

Non-random correspondence of sorghum QTLs from a published database (Zhang et al., 2013) with seven chromosomal 'hotspots' for introgression of sorghum alleles in five geographically diverse US *S. halepense* populations (Morrell et al., 2005) was determined using the hypergeometric probability distribution function, as described (Feltus et al., 2006).

RESULTS

Mosaic Genome of *S. halepense*, With *S. bicolor* Enriched Allele Composition

While its 2.73 ± 0.08 pg/2C genome size closely approximates the sum of those of its progenitors, *S. halepense* has *S. bicolor* enriched allele composition (**Table 1**). To investigate its allele composition, we resequenced tetraploid *S. halepense* accession Gypsum 9 (SRX142088) to a depth of 9.7 Gb, ∼14X coverage of the *S. bicolor* reference genome and conferring ∼95% confidence of detecting *S. halepense* alleles present in as little as one copy. Assuming that tetraploid *S. halepense* has twice the 41,800,275 bp coding DNA sequence (CDS) length of the *S. bicolor* reference genome (Paterson et al., 2009) (**Table 1** and **Supplementary Table 1**), 99.4% of *S. halepense* CDS nucleotides match those of representatives of 'eusorghum (Kellogg, 2013; Hawkins et al., 2015)' progenitor species *S. bicolor* (Paterson et al., 2009) and *S. propinquum* (SRX030701-03), and an outgroup *Sorghum* [*Sarga* (Hawkins et al., 2015)] *timorense* (SRX124552). Among the remaining 500,303 polymorphic nucleotide positions (**Table 1**, patterns 1–15), 10.9% match the *S. bicolor* reference but differ from *S. propinquum* (patterns 2, 3, 8, 9), and 6.6% match *S. propinquum* but not *S. bicolor* (patterns 6, 7, 12, 14). The *S. bicolor* and *S. propinquum* alleles were frequently interleaved along *S. halepense* chromosomes, indicating extensive homogenization (Kong, 2017). This is consistent with largely normal pairing and recombination between *S. bicolor* and *S. propinquum* diploids that is well-known from genetic studies (Chittenden et al., 1994; Paterson et al., 1995; Kong et al., 2015), and with segregation patterns in two interspecific (*S. bicolor* x *S. halepense*) BC_1F_1 populations that suggest a mixture of disomic and polysomic inheritance (Kong, 2017). While our analysis includes an outgroup and compares taxa separated by a minimum of 1–2 million years, some differences among these taxa presumably reflect within-species divergence.

S. halepense Is Richly Polymorphic

Despite a presumed genetic bottleneck during polyploid formation, *S. halepense* is richly polymorphic. A survey of 182 genetically-mapped restriction fragment length polymorphism (RFLP) loci found 18 *S. halepense* or '*Sorghum x almum*' (*S. bicolor* x *S. halepense* hybrid) genotypes to average 6.13 alleles per locus, versus 3.39 for a worldwide sample of 55 landrace and wild sorghum accessions and 1.9 for 16 F1 hybrid sorghums from eight commercial breeding programs (Morrell et al., 2005).

While some apparently novel alleles in the draft genome (**Table 1**) may reflect intraspecific polymorphism, a remarkable 67.1% of CDS polymorphisms differentiate *S. halepense* from representatives of both putative progenitor species and the outgroup *S. timorense* (**Table 1**, pattern 1). The functional impact of these non-synonymous single-nucleotide polymorphisms (SNPs) was assessed by comparison to an evolutionary conservation profile of amino acids from orthologous genes in a panel of diverse plant species, calculating a 'functional impact score' using a modified entropy function (Reva et al., 2011) – 8738 SNPs with high inferred functional impact score' (S_i; see section "Materials and Methods") suggest important consequences for protein function in 5957 *S. halepense* genes (**Supplementary Table 2**). SNPs causing premature protein translation termination (5981 in 4459 genes) are most abundant, followed by loss of stop codons (2521 in 2016 genes) and loss of translation initiation site (236 in 227 genes). These functionally important mutations are significantly enriched in plasma membrane genes with kinase activity, suggesting changes in environmental sensing and associated intracellular processes such as cell differentiation and metabolism (**Supplementary Table 3**).

Rhizomes Are Important to Survival of Both Cold Seasons and Dry Seasons

Rhizomes, subterranean stems that can comprise 70% of its dry weight (Oyer et al., 1959), are a key link between morphology and ecology of *S. halepense*. Rhizome growth of polyploid *S. halepense* transgresses that of its rhizomatous diploid progenitor, *S. propinquum*. We conducted a field trial in Bogart, GA (33.9° N) during 2012-3 of widely spaced (1 m between plants and rows) tetraploid F2 progeny from a cross between *S. bicolor* BTx623 and *S. halepense* Gypsum 9E (*SbxSh*); side by side with plots of 161 diploid recombinant inbred lines from a cross between BTx623 and *S. propinquum* (*SbxSp*; 5 plants per line, spaced 0.3 m between plants and 1 m between rows and plots) (Kong et al., 2015). *SbxSh* progeny had a higher frequency of rhizome-derived shoots emerging from the soil (37.6%), larger average number of rhizomes producing above-ground shoots (0.77), and greater distance of rhizome-derived shoots from the crown (11.97 cm) than *SbxSp* (30%, 0.32, 7.5). Rhizome number showed heritabilities of 0.077 (F_3–F_2 regression) and 0.34 (variance component analysis of BC_1F_2 families) in *SbxSh* and 0.44 in *SbxSp* (by variance component analysis).

Rhizomatousness is closely related to the ability of *S. halepense* to overwinter in the temperate United States. In the Bogart,

TABLE 1 | Coding DNA sequence polymorphism patterns among *Sorghum halepense,* its progenitors *S. propinquum* and wild *S. bicolor*, an elite domesticated *S. bicolor*, and the outgroup *S. timorense* (x indicates sequence divergence, o indicates correspondence).

Pattern	Number	Percentage (excluding pattern 16)	*S. timorense*	*S. propinquum*	*S. bicolor* (wild forms)	*S. bicolor* (reference genome)	Hypothesized interpretation
1	335775	67.1	X	X	X	X	*Novel SH alleles*
2	1259	0.3	X	X	X	O	*Possible introgression from cultivated S. bicolor*
3	1897	0.4	O	X	X	O	*Same as #2*
4	29096	5.8	O	X	X	X	*Alleles reflecting para-eusorghum divergence*
5	1250	0.2	X	O	X	O	*No clear inference*
6	23776	4.8	X	O	X	X	*SP-specific alleles*
7	7160	1.4	O	O	X	X	*Same as #6*
8	37532	7.5	O	X	O	O	*Wild SB-specific alleles*
9	13691	2.7	X	X	O	O	*Same as #8*
10	1652	0.3	O	X	O	X	*Wild SB-specific alleles changed by domestication*
11	5527	1.1	X	X	O	X	*Same as #10*
12	1221	0.2	X	O	O	X	*Ancestral eusorghum alleles changed by domestication*
13	34862	7.0	X	O	O	O	*Ancestral eusorghum alleles*
14	675	0.1	O	O	O	X	*Ancestral para-eusorghum alleles changed by domestication*
15	4930	1.0	O	O	X	O	*No clear inference*
16*	83100247		O	O	O	O	*Conserved ancestral para-eusorghum alleles*
	*Calculation of the number of pattern 16						
	41800275		(a) Total CDS length of *S. bicolor* (improved forms)				
	83600550		(b) Inferred total CDS length of *S. halepense* (2X CDS length of *S. bicolor*)				
	500303		(c) Number of alleles which were not strictly conserved (sum of patterns 1 to 15)				
	83100247		(d) b-c = number of pattern 16				

GA field trial, 139 (58.9% of) *SbxSh* progeny showed regrowth after overwintering, while there was no survival of *SbxSp* in 2012-3 or in two additional years. Moreover, in *SbxSh* BC_1F_1-derived BC_1F_2 families (n = 246) grown in 3 m plots with two replications following conventional sorghum recommendations, those with rhizomes had significantly higher frequencies of survival than those lacking rhizomes (**Table 2**). The advantage of rhizomes was observed both in harsh winters (2013-14, with five periods below 20 F, reaching a low of 5.8 F[1]) and mild winters (2014-15, with only two periods below 20 F, reaching a low of 10.2 F) in Bogart GA. Survival in Salina, KS among replica plots of the same BC_1F_2 families was too low to evaluate statistically.

TABLE 2 | Overwintering of *S. bicolor x S. halepense* BC_1F_1-derived BC_1F_2 families is related to rhizomatousness.

	Rhizomes	No rhizomes
a: 2013-4, Bogart, GA, χ^2 = 8.84, 1 d.f., p = 0.001.		
Survived	24	1
Died	159	77
b: 2014-5, Bogart, GA, χ^2 = 3.10, 1 d.f., p = 0.08.		
Survived	116	28
Died	76	31

More extensive rhizome growth than its rhizomatous diploid progenitor is also related to the ability of *S. halepense* to survive tropical dry seasons. From a total of 96 BC_1F_2 families selected for rhizome growth in Bogart GA, single 3 m rows were tested for 15 months (2014-5) at the ICRISAT research station in Samanko, Mali (12.5° N, −7.9° W). A total of 45 (47% of) families contained one or more plants that survived the dry season of 8 month duration with zero rainfall. A logistic regression model (see section "Materials and Methods") showed that for each 1 cm increase in rhizome spread from the crown based on Bogart GA trials, the probability of surviving the Malian dry season increased ∼3%. Factors other than rhizomes are also important to perenniality – lines surviving the tropical dry

[1]http://www.georgiaweather.net/index.php?variable=HI&site=WATUGA

season were only randomly associated with those surviving the mild 2014-15 temperate winter in Bogart, GA (24 of 54 lines, 44%), survivors of the harsh 2013-14 winter being more closely associated with dry season survival but too few in overall number to be conclusive (5 of 6, 83%).

Rhizome Growth Is Correlated With Reproduction

Curiously, rhizome growth is correlated negatively with that of other vegetative organs but positively with reproductive growth. Across four environments (Bogart GA and Salina KS, 2013 and 2014), early flowering is correlated with reduced aboveground vegetative biomass ($r = -0.26$ to -0.62, $p < 0.001$), but increased rhizome growth ($r = 0.17$ to 0.30, $p < 0.001$) in tetraploid *SbxSh* progeny. Because rhizomes are a vegetative organ, our *a priori* expectation was that increased vegetative biomass aboveground would be correlated with increased rhizome growth. However, we measured rhizome growth primarily based on counting aboveground shoots derived from rhizomes. In another rhizomatous grass (*Agropyron repens*), rhizome axillary buds experience apical dominance until anthesis, being suppressed by auxins (Leakey et al., 1975). By excising *S. halepense* rhizomes from the plant, we found that axillary buds consistently develop as vertical shoots and not as rhizomes (**Figure 1**). So, once flowering of the primary stalk is initiated, a rhizomatous plant permits the development of additional ramets – which in principle should be able to exert apical dominance themselves. Moreover, our observation that these new buds invariably become ramets and not rhizomes raises questions about their additional dependence on a mobile 'florigen' such as that translocated to the plant apex (Sachs, 1865). There may be much to be learned about nature of signaling among ramets at different developmental stages that are interconnected by rhizomes.

Both Polyploidy and Interspecific Hybridity Appear to Contribute to the 'Mosaic' Nature of Rhizome Gene Expression

While ∼80% of annotated sorghum genes are expressed in *S. halepense* rhizomes, many alleles with striking enrichment ($p < 0.001$) of expression more closely resemble the sequences of the non-rhizomatous *S. bicolor* progenitor than rhizomatous *Sp*. By laser capture microdissection, we collected meristems and compared transcripts from buds induced to develop as rhizomes or leafy shoots (**Figure 1**), respectively obtaining 163,264,254 and 152,162,240 Illumina Hiseq reads, of which 67.7% (110,492,577) and 67.2% (102,194,352) could be anchored to 27,566 and 27,183 sorghum gene models. About 1% (262) of genes showed differential expression ($p < 0.001$) between rhizome buds (168 enriched) and shoot buds (94: **Supplementary Table 4**). Appreciable recruitment of alleles from non-rhizomatous *S. bicolor* to rhizome-enriched expression is indicated by 44 *S. bicolor* versus only 23 *S. propinquum* derived transcripts with at least two SNPs supporting these origins and no contradictory SNPs (other differentially expressed genes are ambiguous based on these criteria).

Consistent with rhizomes being ∼70% of the mass of a Johnsongrass plant (Oyer et al., 1959), genes highly expressed in rhizome buds were enriched for diverse functions associated with rapid cell division (Kinesins, ATP binding, and microtubule related: **Supplementary Table 5**) and maturation. Cellulose synthase, Sb06g016760, was the most rhizome enriched gene, also implicated in rapid cell growth. Shoot-bud enriched genes were over-represented in three gene ontology (GO) categories associated with cell recognition (**Supplementary Table 5**), perhaps in preparation for new biotic interactions after emergence from the soil. The most shoot-enriched genes were (a) glutathione *S*-transferase (Sb09g000860), catalyzing conjugation of the reduced form of glutathione (GSH) to xenobiotic substrates for detoxification; (b) a glycoside hydrolase (Sb08g007610), suggesting cell wall loosening during the rhizome-to-shoot transition; and (c) a member of the major facilitator superfamily (Sb06g033080, MFS: Interpro IPR005828) of transmembrane single-polypeptide secondary carriers implicated in control of sorghum seed size (Zhang et al., 2015), a trait that shows strong negative correlation with both rhizome development and winter survival (TSC, personal communication). Intriguing differentially expressed genes located within likelihood intervals of rhizome related quantitative trait loci (QTLs, **Figure 2**) include an auxilin/cyclin G-associated kinase (Sb03g028900), tandemly duplicated ethylene responsive transcription factors (Sb07g006195, Sb07g006200), and a Ca2 + /calmodulin-dependent protein kinase, EF-Hand protein superfamily gene (Sb09g022960).

Both polyploidy and interspecific hybridity appear to contribute to the 'mosaic' nature of rhizome gene expression, with overexpression of some homoeologs from rhizomatous *S. propinquum* and others from non-rhizomatous *S. bicolor* (**Supplementary Table 4**). For example, different calmodulin family members have evolved specificity to rhizome buds (e.g., Sb10g027610, the second-most rhizome specific gene) and shoot buds (Sb06g023700). Tandem duplicated ethylene responsive transcription factors within a rhizome-related QTL are both overexpressed in *S. halepense* rhizome buds, although the sequence of Sb07g006195 closely resembles *S. propinquum* (5 of 6 SNPs matching) and adjacent Sb07g006200 is identical to *S. bicolor* (6 of 6 SNPs). The *Teosinte-branched 1* growth repressor gene implicated in apical dominance of maize shoots has two family members with enriched expression in rhizome buds (Sb01g010690, Sb04g026970), ironically both completely matching the non-rhizomatous *S. bicolor* progenitor sequences (4 of 4, and 2 of 2 SNPs).

Adaptation by *S. halepense* to New Continents and Latitudes May Have Been Facilitated by Introgression From Cultivated Sorghum

Introgression is suggested in a general sense by *S. bicolor* enriched allele composition of the *S. halepense* draft genome (**Table 1**), and for specific genes by *S. halepense* SNP distribution patterns matching the *S. bicolor* reference genome of an elite breeding line (Paterson et al., 2009), but differing from both several

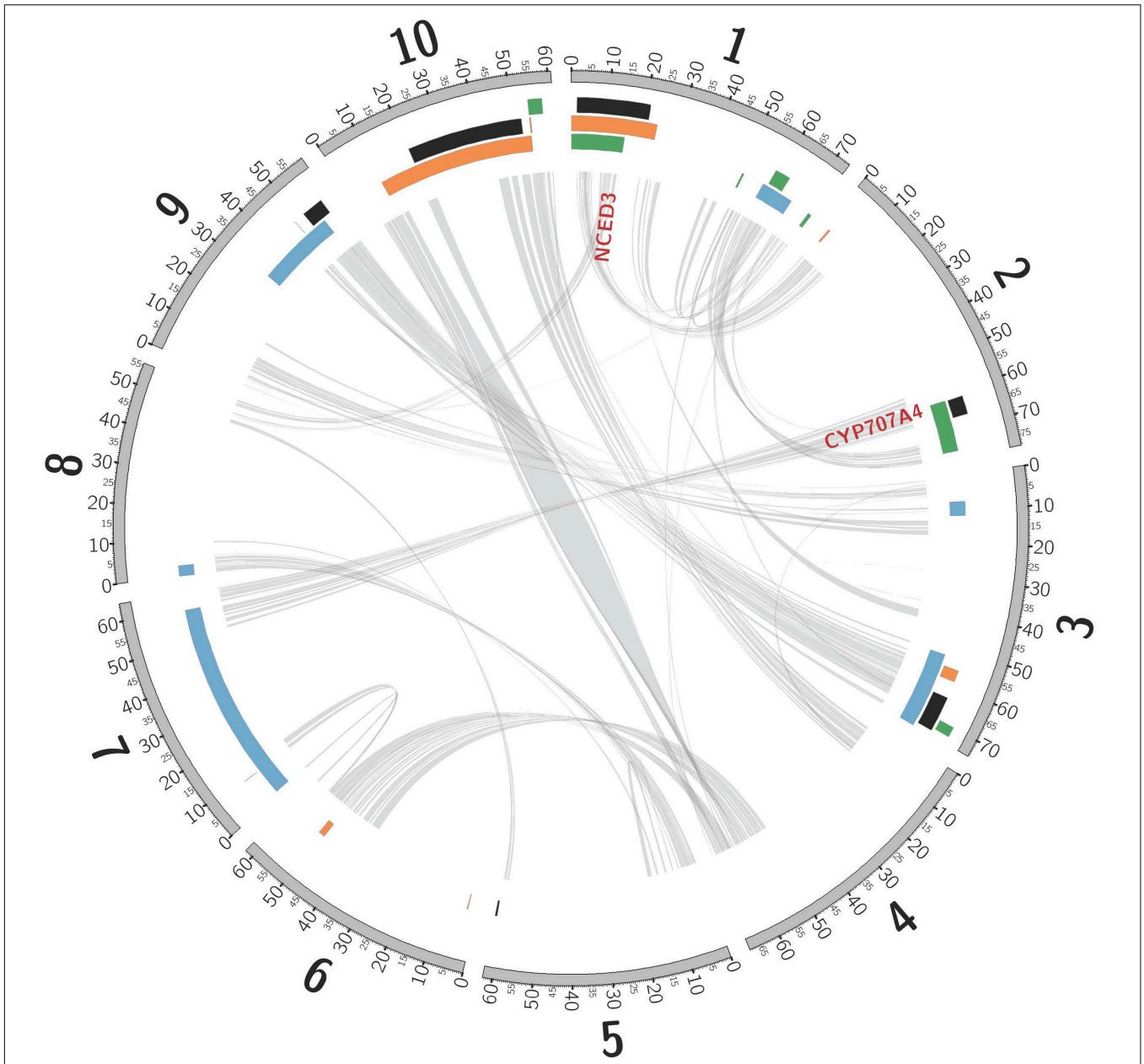


FIGURE 2 | Non-random association between QTLs mapped in sorghum, and introgression from *S. bicolor* into *S. halepense*. Sorghum chromosomes (units are megabases), annotated with physical locations of hotspots of introgression from *S. bicolor* into *S. halepense* [black (Morrell et al., 2005)] and QTLs for seed lutein concentration [green (Fernandez et al., 2008)], seedling vigor [orange (Fernandez et al., 2008)], and rhizomes [blue (Zhang et al., 2013)]. NCED3 and CYP707A4 are gene candidates for cold tolerance (see text). Internal lines indicate syntenic duplicated genes persisting from whole-genome duplication ~96 million years ago (Paterson et al., 2009; Wang et al., 2015).

wild *S. bicolors* and each of two outgroups (**Table 1**, patterns 2–3). Seven 'hotspots' for introgression of sorghum alleles in five geographically diverse US *S. halepense* populations (Morrell et al., 2005), show non-random correspondence with published sorghum QTLs (Zhang et al., 2013) conferring variation in rhizome growth, seed size, and lutein content (**Figure 2** and **Table 3**). While sorghum lacks rhizomes and has large seeds, rhizome growth-related alleles masked in domesticated sorghum genotypes by a lack of rhizomes may be unmasked in interspecific crosses with rhizomatous *S. halepense*.

Particularly intriguing among *S. halepense* introgression hotspots are those that correspond with 3 of 4 QTL likelihood intervals spanning 4.9% of the genome that account for variation in seed content of the carotenoid lutein (Fernandez et al., 2008) (p = 0.0026, **Table 3**). Sorghum leaf photosynthetic capacity is susceptible to damage under low-temperature (<10

TABLE 3 | Sorghum QTLs associated with chromosomal 'hotspots' for introgression of sorghum alleles into five geographically diverse *S. halepense* populations (Morrell et al., 2005).

Trait	Hotspots	% of genome	# QTLs	Enrichment	*p*-Value (hypergeometric)
Rhizome number	4	13.3	8	4.30	8.82E-05
Seed size	4	8.6	4	6.64	5.14E-04
Lutein	3	5.0	4	8.64	2.58E-03

C) but high-light conditions when electron transport exceeds the capacity of carbon fixation to utilize available energy (Taylor and Rowley, 1971). Such conditions are infrequent in the tropics where *Sorghum* originated but common in the temperate springtime. Spring regrowth of *S. halepense* starts about 4 weeks before cultivated sorghum is seeded at 38.7° N (Gypsum, KS, where Gypsum 9 was collected). Xanthophyll carotenoids such as lutein are most abundant in plant leaves, modulating light energy and performing non-photochemical quenching of excited 'triplet' chlorophyll which is overproduced at very high light levels during photosynthesis (Demmig-Adams and Adams, 2006; Taiz and Zeiger, 2006). Ironically, Sb01g030050 (*Lut1*; KO:K09837) and Sb01g048860 (*crtZ*; KO:K15746) related to lutein biosynthesis, are close to the only lutein QTL not near an introgression hotspot (on chromosome 1).

Within the lutein QTL likelihood intervals, and homozygous in the Gypsum 9E (**Supplementary Table 6**), are also loss of function mutations in Sb01g013520, 9-*cis* epoxycarotenoid dioxygenase. This enzyme cleaves xanthophylls to xanthoxin, a precursor of the plant hormone abscisic acid (ABA) (Tan et al., 2003) that plays a central role in regulating plant tissue quiescence. Also in the lutein QTL likelihood intervals are non-synonymous SNPs inferred to have striking functional effects (see section "Materials and Methods") on Sb02g026600, a cytochrome P450 performing a key step of ABA catabolism (Saito et al., 2004). A hypothesis for investigation is whether modified alleles at these loci degrade ABA to release *S. halepense* seeds from dormancy early and/or increase seedling vigor under cold conditions.

DISCUSSION

Synergy between gene duplication and interspecific hybridity may add an important element to the classical notion that polyploids adapt better than their diploid progenitors to environmental extremes (Muntzing, 1936; Love and Love, 1949; Stebbins, 1950; Grant, 1971). Evidence is growing that polyploidy is an important contributor to biological invasions (te Beest et al., 2012). Genome duplication facilitates the evolution of genes with new or modified functions (Ohno, 1970) such as we report, permitting a nascent polyploid to adapt to environments beyond the reach of its progenitors. Hybridity preserves novel alleles such as many recruited into *S. halepense* rhizome-enriched gene expression from non-rhizomatous *S. bicolor*, putatively contributing to the transgressive rhizome growth and ability of *S. halepense* but not rhizomatous *S. propinquum* derived progeny to overwinter in the temperate United States.

Several lines of evidence point to a richness of DNA-level variation in *S. halepense*, including an abundance of novel coding sequences, much richer diversity of neutral DNA markers than its progenitors, and novel gene expression patterns exemplified by rhizome-enriched expression of some alleles from its non-rhizomatous *S. bicolor* progenitor. The spread of invasive taxa is much more rapid than migration in native taxa, and may require more genetic variation to sustain (Lee, 2002). Although there is somewhat less variation near the invasion front than the center of its US distribution (Sezen et al., 2016), rich *S. halepense* diversity may support its projected 200–600 km northward spread in the coming century (McDonald et al., 2009).

Rich genetic variation in *S. halepense* offers not only challenges but also opportunities. Long under selection for weediness-related attributes that enhance its competitiveness with crops, some US *S. halepense* genotypes have transitioned to non-agricultural niches (Sezen et al., 2016) and may also experience selection favoring alleles that could improve sorghum and other crops, e.g., for cold tolerance, rapid vegetative development and flowering, disease and pest resistance, and ratooning (a new growth cycle from the stubble of the prior one). *Sorghum bicolor* can routinely serve as the pollen parent of triploid and tetraploid (reviewed in Warwick and Black, 1983; Tang and Liang, 1988) and under some circumstances diploid (Dweikat, 2005; Cox et al., 2017), interspecific hybrids with *Sh*, offering the opportunity to test *S. halepense* alleles in sorghum.

As the first surviving polyploid in its lineage in ∼96 million years (Paterson et al., 2009; Wang et al., 2015), *S. halepense* may open new doors to sorghum improvement, with synergy between gene duplication and interspecific hybridity nurturing the evolution of genes with new or modified functions (Ohno, 1970). Already, genetic novelty from *S. halepense* is being used in efforts to breed ratooning/perennial sorghums that better protect 'ecological capital' such as topsoil and organic matter (Glover et al., 2010). Attributes of *S. halepense* such as endophytic nitrogen fixation (Rout et al., 2013), if transferred to sorghum, could help to narrow a 'yield gap' reflected by 1961–2012 yield gains in the U.S. of only 61% for sorghum versus 323% for maize[2]. Likewise, its perenniality may have resulted in selection for 'durable' biotic stress resistance mechanisms that are absent from, but of importance to the improvement of, sorghum and other crops.

DATA AVAILABILITY STATEMENT

The dataset(s) used in this study can be found as follows: the resequenced genome for tetraploid *S. halepense* accession Gypsum 9 is archived under NCBI ID SRX142088.

[2]http://faostat.fao.org

The resequenced genome for *S. propinquum* is archived under NCBI ID SRX030701-03. The resequenced genome for *Sorghum* [*Sarga* (Hawkins et al., 2015)] *timorense* is archived under NCBI ID SRX124552. LM RNA-seq reads are archived under NCBI BioProject ID PRJNA356741.

AUTHOR CONTRIBUTIONS

AP contributed conception and design of the study and wrote the first draft of the manuscript. WK, RJ, PN, VG, KI, US, MK, DB, EW, HR, JB, KB, TC, and MS collected field and/or laboratory data. GW, WP, T-HL, HG, DZ, and FF performed statistical analyses. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

We appreciate support for aspects of this work from the NIFA Global Food Security CAP (2015-68004-23492 to AP, JB), USAID Feed The Future (AID-OAA-A-13-00044 to AP, TC, EW, HR), US Department of Energy Joint Genome Institute Community Sequencing Program (to AP), AFRI Plant Growth and Development Program (2009-03477 and 2016-67013-24608 to AP and MS), USDA Biotechnology Risk Assessment Program (2012-01658 to AP and TC), and AFRI Controlling Weedy and Invasive Plants Program (2013-67013-21306 to JB and AP). The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231.

ACKNOWLEDGMENTS

We thank Daniel S. Rokhsar for assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00317/full#supplementary-material

REFERENCES

Arumuganathan, K., and Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9, 208–218. doi: 10.1007/bf02672069

Binimelis, R., Pengue, W., and Monterroso, I. (2009). Transgenic treadmill: responses to the emergence and spread of glyphosate-resistant *Johnsongrass* in argentina. *Geoforum* 40, 623–633. doi: 10.1016/j.geoforum.2009.03.009

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Celarier, R. (1958). Cytotaxonomic notes on the subsection *Halepense* of the genus Sorghum. *Bull. Torrey Bot. Club.* 85, 49–62.

Chittenden, L. M., Schertz, K. F., Lin, Y. R., Wing, R. A., and Paterson, A. H. (1994). A detailed RFLP map of sorghum bicolor X S. propinquum, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. *Theor. Appl. Genet.* 87, 925–933. doi: 10.1007/bf00225786

Cox, S., Nabukalu, P., Paterson, A. H., Kong, W., Auckland, S., Rainville, L., et al. (2017). High proportion of diploid hybrids produced by interspecific diploid × tetraploid *sorghum* hybridization. *Genet. Resourc. Crop Evol.* 65, 387–390. doi: 10.1007/s10722-017-0580-7

da Huang, W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., et al. (2007). The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8:R183.

Demmig-Adams, B., and Adams, W. W. (2006). dams: photoprotection in an ecological context: the remarkable complexity of thermal energy dissipation. *New Phytol.* 172, 11–21. doi: 10.1111/j.1469-8137.2006.01835.x

Doggett, H. (1976). "Sorghum," in *Evolution of Crop Plants*, ed. N. Simmonds (Essex: Longman).

Dweikat, I. (2005). A diploid, interspecific, fertile hybrid from cultivated sorghum. sorghum bicolor, and the common *Johnsongrass* weed *Sorghum halepense*. *Mol. Breed.* 16, 93–101. doi: 10.1007/s11032-005-5021-1

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Brown, P., Klein, P. E., et al. (2006). Genetic map alignment and QTL correspondence between inter- and intra-specific sorghum populations. *Theor. Appl. Genet.* 112, 1295–1305. doi: 10.1007/s00122-006-0232-3

Feltus, F. A., Wan, J., Schulze, S. R., Estill, J. C., Jiang, N., and Paterson, A. H. (2004). An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. *Genome Res.* 14, 1812–1819. doi: 10.1101/gr.2479404

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. P., and Kresovich, S. (2008). Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. *Crop Sci.* 48, 1732–1743. doi: 10.2135/cropsci2007.12.0684

Glover, J. D., Reganold, J. P., Bell, L. W., Borevitz, J., Brummer, E. C., Buckler, E. S., et al. (2010). Increased food and ecosystem security via perennial grains. *Science* 328, 1638–1639. doi: 10.1126/science.1188761

Grant, V. (1971). *Plant Speciation*, 1st Edn, New York, NY: Columbia University Press.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nrot.2013.084

Hawkins, J. S., Ramachandran, D., Henderson, A., Freeman, J., Carlise, M., Harris, A., et al. (2015). Phylogenetic reconstruction using four low-copy nuclear loci strongly supports a polyphyletic origin of the genus Sorghum. *Ann. Bot.* 116, 291–299. doi: 10.1093/aob/mcv097

Heap, I. (2012). *The International Survey of Herbicide Resistant Weeds*. Ames, IA: Iowa State University Extension and Outreach.

Kellogg, E. (2013). "Phylogenetic relationships of Saccharinae and Sorghinae," in *Genomics of the Saccharinae*. ed. A. H. Paterson (New York, NY: Springer).

Kong, W. (2017). "Genetic dissection of plant architecture and life history traits salient to climate-resilient sustainable intensification of agriculture," in *Department of Crop and Soil Science* (Athens, GA: University of Georgia).

Kong, W., Kim, C., Goff, V. H., Zhang, D., and Paterson, A. H. (2015). Genetic analysis of rhizomatousness and its relationship with vegetative branching of Sorghum bicolor × S. propinquum recombinant inbred lines. *Am. J. Bot.* 102, 718–724. doi: 10.3732/ajb.1500035

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., and McWilliam, H. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Leakey, R. R. B., Chancellor, R. J., and Vince-Prue, D. (1975). Parental factors in dominance of lateral buds on rhizomes of *Agropyron repens* (L.). *Beauv. Plant.* 123, 267–274. doi: 10.1007/BF00390705

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends Ecol. Evol.* 17, 386–391. doi: 10.1016/s0169-5347(02)02554-5

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Love, A., and Love, D. (1949). The geobotanical significance of polyploidy. *Portug. Acta* 5, 273–352.

McDonald, A., Riha, S., DiTommaso, A., and DeGaetano, A. (2009). Climate change and the geography of weed damage: analysis of US maize systems suggests the potential for significant range transformations. *Agric. Ecosyst. Environ.* 130, 131–140. doi: 10.1016/j.agee.2008.12.007

McWhorter, C. G. (1971). Introduction and spread of *Johnsongrass* in the United States. *Weed Sci.* 19:496. doi: 10.1017/s0043174500050517

Morrell, P. L., Williams-Coplin, D., Bowers, J. E., Chandler, J. M., and Paterson, A. H. (2005). Crop-to-weed introgression has impacted allelic composition of *Johnsongrass* populations with and without recent exposure to cultivated sorghum. *Mol. Ecol.* 14, 2143–2154. doi: 10.1111/j.1365-294x.2005.02579.x

Muntzing, A. (1936). The evolutionary significance of autopolyploidy. *Hereditas* 21, 363–378. doi: 10.1111/j.1601-5223.1936.tb03204.x

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer.

Oyer, E., Gries, G., and Rogers, B. (1959). The seasonal reproduction of johnson grass plants. *Weeds* 7:13. doi: 10.2307/4040251

Paterson, A., Schertz, K., Lin, Y., Liu, S., and Chang, Y. (1995). The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of *Johnsongrass*. *Sorghum halepense* (L.). *Pers. Proc. Natl. Acad. Sci. U.S.A.* 92, 6127–6131. doi: 10.1073/pnas.92.13.6127

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., and Gundlach, H. (2009). The sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723

Quinn, L., Barney, J. N., McCubbins, J., and Endres, A. (2013). Navigating the "Noxious" and "Invasive". Regulatory landscape: suggestions for improved regulation. *Bioscience* 63, 124–131. doi: 10.1525/bio.2013.63.2.8

Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39:e118. doi: 10.1093/nar/gkr407

Rout, M. E., Chrzanowski, T. H., DeLuca, T. H., Westlie, T. K., Callaway, R. M., and Holben, W. E. (2013). Bacterial endophytes enhance invasive plant competition. *Am. J. Bot.* 100, 1726–1737. doi: 10.3732/ajb.1200577

Sachs, J. (1865). Wirkung des lichtes auf die blütenbildung unter vermittlung der laubblätter. *Bot. Ztg.* 23, 117–121.

Saito, S., Hirai, N., Matsumoto, C., Ohigashi, H., Ohta, D., Sakata, K., et al. (2004). Arabidopsis CYP707As encode (+)-abscisic acid 8′-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. *Plany Physiol.* 134, 1439–1449. doi: 10.1104/pp.103.037614

Sezen, U. U., Barney, J. N., Atwater, D. Z., Pederson, G. A., Pedersen, J. F., Chandler, J. M., et al. (2016). Multi-phase US spread and habitat expansion of a post-columbian invasive. *Sorghum halepense*. *PLoS One* 11:e01644584. doi: 10.1371/journal.pone.0164584

Stebbins, G. L. (1950). *Variation and Evolution in Plants*. New York, NY: Columbia University Press.

Taiz, L., and Zeiger, E. (2006). *Plant Physiology*. Sunderland, MA: Sinauer Associates Inc.

Takacs, E. M., Li, J., Du, C., Ponnala, L., Janick-Buckner, D., Yu, J., et al. (2012). Ontogeny of the maize shoot apical meristem. *Plant Cell* 24, 3219–3234.

Tan, B. C., Joseph, L. M., Deng, W. T., Liu, L., Li, Q. B., Cline, K., et al. (2003). Molecular characterization of the *Arabidopsis* 9-cis epoxycarotenoid dioxygenase gene family. *Plant J.* 35, 44–56.

Tang, H., and Liang, G. H. (1988). The genomic telationship between cultivated sorghum Sorghum bicolor (L) Moench and *johnsongrass* [*Sorghum halepense* (L) Pers] - a reevaluation. *Theor. Appl. Genet.* 76, 277–284. doi: 10.1007/bf00257856

Taylor, A. O., and Rowley, J. A. (1971). Plants under climatic stress: I. Low temperature, high light effects on photosynthesis. *Plant Physiol.* 47, 713–718. doi: 10.1104/pp.47.5.713

te Beest, M., Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., and Kubešova, M. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot.* 109, 19–45. doi: 10.1093/aob/mcr277

Tellman, B. (ed.) (1996). "Stowaways and invited guests: how some exotic plants reached the american southwest," in *California Exotic Pest Council 1996 Symposium*, (San Diego: California Exotic Pest Council).

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Wang, X., Wang, J., Guo, H., Jin, D., Lee, T.-H., Liu, T., et al. (2015). Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Mol. Plant* 8, 885–898. doi: 10.1016/j.molp.2015.04.004

Warwick, S. I., and Black, L. D. (1983). The biology of canadian weeds. 61. *Sorghum halepense* (L.). *Pers. Can. J. Plant Sci.* 63, 997–1014. doi: 10.4141/cjps83-125

Zhang, D., Guo, H., Kim, C., Lee, T. H., Li, J., Robertson, J., et al. (2013). H.: CSGRqtl, a comparative QTL database for saccharinae grasses. *Plant Physiol.* 161, 594–599. doi: 10.1104/pp.112.206870

Zhang, D., Li, J., Compton, R. O., Robertson, J., Goff, V. H., Epps, E., et al. (2015). Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (Panicoidae) and rice (Oryzoidae). *G3* 3, 1117–1128. doi: 10.1534/g3.115.017590

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 July 2020
doi: 10.3389/fgene.2020.00687

Balanced Genome Triplication in Wheat Causes Premature Growth Arrest and an Upheaval of Genome-Wide Gene Regulation

*Xiaowan Gou[1,2], Ruili Lv[2], Changyi Wang[2], Tiansi Fu[2], Yan Sha[2], Lei Gong[2], Huakun Zhang[2] and Bao Liu[2]**

[1] School of Life Sciences, Jiangsu Normal University, Xuzhou, China, [2] Key Laboratory of Molecular Epigenetics, Northeast Normal University, Changchun, China

Polyploidy, or whole genome duplication (WGD), is a driving evolutionary force across the tree of life and has played a pervasive role in the evolution of the plant kingdom. It is generally believed that a major genetic attribute contributing to the success of polyploidy is increased gene and genome dosage. The evolution of polyploid wheat has lent support to this scenario. Wheat has evolved at three ploidal levels: diploidy, tetraploidy, and hexaploidy. Ample evidence testifies that the evolutionary success, be it with respect to evolvability, natural adaptability, or domestication has dramatically increased with each elevation of the ploidal levels. A long-standing question is what would be the outcome if a further elevation of ploidy is superimposed on hexaploid wheat? Here, we characterized a spontaneously occurring nonaploid wheat individual in selfed progenies of synthetic hexaploid wheat and compared it with its isogenic hexaploid siblings at the phenotypic, cytological, and genome-wide gene-expression levels. The nonaploid manifested severe defects in growth and development, albeit with a balanced triplication of the three wheat subgenomes. Transcriptomic profiling of the second leaf of nonaploid, taken at a stage when phenotypic abnormality was not yet discernible, already revealed significant dysregulation in global-scale gene expression with *ca.* 25.2% of the 49,436 expressed genes being differentially expressed genes (DEGs) at a twofold change cutoff relative to the hexaploid counterpart. Both up- and downregulated DEGs were identified in the nonaploid vs. hexaploid, including 457 genes showing qualitative alteration, i.e., silencing or activation. Impaired functionality at both cellular and organismal levels was inferred from gene ontology analysis of the DEGs. Homoeologous expression analysis of 9,574 sets of syntenic triads indicated that, compared with hexaploid, the proportions showing various homeologous expression patterns were highly conserved in the nonaploid although gene identity showed moderate reshuffling among some of the patterns in the nonaploid. Together, our results suggest hexaploidy is likely the upper limit of ploidy level in wheat; crossing this threshold incurs severe ploidy syndrome that is preceded by disruptive dysregulation of global gene expression.

Keywords: ploidy level, genome multiplication, transcriptome shock, dysregulation, ploidy syndrome, *Triticum aestivum*

OPEN ACCESS

Edited by:
Yves Van de Peer,
Ghent University, Belgium

Reviewed by:
Guanjing Hu,
Iowa State University, United States
Jianghua Chen,
Xishuangbanna Tropical Botanical Garden (CAS), China

****Correspondence:***
Bao Liu
baoliu@nenu.edu.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *07 April 2020*
Accepted: *04 June 2020*
Published: *08 July 2020*

Citation:
Gou X, Lv R, Wang C, Fu T, Sha Y, Gong L, Zhang H and Liu B (2020) Balanced Genome Triplication in Wheat Causes Premature Growth Arrest and an Upheaval of Genome-Wide Gene Regulation. Front. Genet. 11:687. doi: 10.3389/fgene.2020.00687

INTRODUCTION

Polyploidy, or whole-genome duplication (WGD), is an important evolutionary force that has shaped the genomes of all higher plants and many animals (Van de Peer et al., 2009; Jiao et al., 2011; Paterson et al., 2012; Soltis and Soltis, 2012; Jiao and Paterson, 2014; Wendel, 2015; Schwager et al., 2017; Van de Peer et al., 2017; Blischak et al., 2018; Spoelhof et al., 2019). Apart from evolutionary importance, polyploidy also bears significant contemporary relevance to human health (Gjelsvik et al., 2019), and crop improvement and is becoming increasingly topical in diverse research fields. Myriad attributes characterize polyploidy relative to its diploid counterpart with genome-wide increase of DNA content and gene dosage being at the root. Interspecific hybridization is another major evolutionary force that facilitates adaptation and speciation primarily via new combinations of old parental genetic variants (Soltis and Soltis, 2009; Marques et al., 2019), but it may also include immediate effects of genome shock (McClintock, 1984) that induce new genetic and epigenetic variations (Kashkush et al., 2002; Senerchia et al., 2015). Interspecific hybridization either occurs at the homoploid level or is coupled with WGD, i.e., allopolyploidization, with the latter being particularly prevalent in plants. The combined effects of genome merger and WGD may cause a stronger genome shock and are more potent in the generation of heritable variations than WGD alone as evidenced by a large body of empirical studies (Adams, 2007; Doyle et al., 2008; Feldman and Levy, 2012; Madlung and Wendel, 2013; Yoo et al., 2014; Song and Chen, 2015; Wendel, 2015). Consequently, polyploidy in general and allopolyploidy in particular are often associated with enhanced organismal performance in growth, yield, fitness, and evolvability (Comai, 2005; Doyle et al., 2008; Wendel, 2015; Van de Peer et al., 2017).

Apart from genetic consequences of WGD, many aspects of physiology are instantaneously affected in polyploids, which contribute to their altered metabolism and phenotypes, such as nuclear volume; cell size and number; and organ/organism structure, shape, and size. In fact, every aspect of cell activity can be affected in a polyploid due to physiological alterations (Doyle and Coate, 2019) because different cellular components may not scale proportionally with elevated ploidy level (Storchova et al., 2006). Thus, altered physiology alone without invoking genetic properties may constitute a constraint limiting the extent to which the number of genome multiplications can be tolerated by a given organism. Notably, while the extent of genome multiplication in certain terminally differentiated tissues/organs does not appear to be under strict control (Doyle and Coate, 2019), it apparently does at the organismal level. Indeed, several studies have shown that there exists an upper limit for the number of genome multiplications at the organismal level for a given species. For instance, Corneillie et al. (2019) assessed growth rates, biomass, and cell wall composition in *Arabidopsis thaliana* autopolyploid plants of three ploidal levels ($2n$ = $4x$, $6x$, and $8x$) relative to their isogenic diploid ($2x$) counterpart and found that only tetraploid exhibited superior performance in all the analyzed traits, while hexaploid and octaploid plants did not show advantages in all traits or even manifested apparent defects in most traits in the octaploid. This is consistent with earlier findings in *A. thaliana* by Tsukaya (2008), who described the phenomenon "high ploidy syndrome," i.e., higher ploidy level often exhibits retarded growth and trade-offs between cellular and organ size or organismal size. The detrimental effects on growth by higher ploidy are thought to be due to burdens on cell cycle and, hence, reduced cell division rate as a result of increased DNA content (Tsukaya, 2008).

Conceivably, interplays exist between the physiological effects and genetic consequences of genome multiplication. For example, disproportional scaling in surface area and length of spindle pole body generates geometric constraints that results in chromosomal instability in yeast (Storchova et al., 2006). Also, genome-wide changes in gene expression are expected to be an automatic outcome, i.e., transcriptomic response, as a result of a genome-wide increase of gene dosage. Although gene expression analyses associated with polyploidization have been mainly conducted in allopolyploids in which the combined effects of hybridization and WGD are involved (Doyle et al., 2008; Yoo et al., 2014; Song and Chen, 2015), several recent studies have focused on autopolyploids, i.e., the pure effect of genome multiplication. It was found that WGD *per se* alters expression of a substantial proportion of the transcriptome, especially when changes in transcriptome size was taken into account (Robinson et al., 2018; Doyle and Coate, 2019; Spoelhof et al., 2019; Visger et al., 2019). Hitherto, a relationship between transcriptomic response and "high ploidy syndrome" (Tsukaya, 2008) remains poorly understood.

The wheat genus (*Triticum*) comprises species at three ploidal levels ($2x$, $4x$, and $6x$), and remarkably, at least one species at each ploidal level has been successfully domesticated into major food crops (Matsuoka, 2011; Feldman and Levy, 2015). Nevertheless, wheat species of the three ploidal levels differ dramatically with respect to both their evolvability under natural conditions, for example, differential niche expansion ranges between wild diploid and tetraploid wheat (Salamini et al., 2002), and their human-mediated dispersions, which are dramatically different among the domesticated diploid, tetraploid, and hexaploid wheat (Dubcovsky and Dvorak, 2007). In fact, the emergence of each domesticated wheat of a higher ploidy (tetraploid vs. diploid and hexaploid vs. tetraploid) has rapidly replaced the lower ploidy one. This suggests superiority of increased ploidal level in the wheat species as allopolyploids. A long-standing issue, which to our knowledge has not been addressed, is what would be the outcome if a further whole genome multiplication is superimposed on hexaploid wheat?

Here, we characterize a spontaneously occurring nonaploid wheat individual in selfed progenies of a synthetic hexaploid wheat and compared it with its isogenic hexaploid siblings at the phenotypic, cytological, and genome-wide gene-expression levels. Our results suggest triplication of the three subgenomes of hexaploid wheat, let alone expected problems in meiosis as anisoploidy already imposes a severe detrimental effect on global gene regulation and growth/development at the

vegetative stages, suggesting hexaploidy is likely the upper limit of ploidal level in wheat.

MATERIALS AND METHODS

Plant Materials

The seeds of allohexaploid wheat line AT5 at the S_0 generation along with its parental lines were procured from Moshe Feldman (Weizmann Institute of Science, Israel). This line was produced by hybridization of *Triticum turgidum* ssp. *durum* (accession TTR04) and *Aegilops tauschii* (accession TQ27), followed by colchicine treatment to induce chromosome doubling. All plant individuals were propagated under strict selfing conditions to produce advanced generations of AT5 from S_0 to S_6. All plants were grown in a common condition (day/night, 25°C/16°C, 16 h/8 h). When the third leaf appeared, and the second leaf fully expanded (Simmons et al., 1985), the second leaf of hexaploidy and nonaploidy was collected for RNA isolation. All collected leaves were kept at −80°C until use.

Karyotyping by Sequential Fluorescence *in situ* Hybridization and Genomic *in situ* Hybridization

The protocols for fluorescence *in situ* hybridization (FISH) and genomic *in situ* hybridization (GISH) were essentially as described in Kato et al. (2004) with minor modifications. For FISH, two repetitive DNA sequences (pSc119.2 and pAS1) were labeled by nick translation with Alexa Fluor 488-5-dUTP (green) and Texas Red-5-dCTP (red), respectively. For GISH, genomic DNA of *T. urartu* and *Ae. tauschii* was labeled with Alexa Fluor 488-5-dUTP (green coloration), and Texas Red-5-dCTP (red coloration), respectively. Genomic DNA of *Ae. speltoides* was used as a blocker.

Metaphase chromosome spreads were prepared following Kato et al. (2004). Slide denaturation, hybridization, and washing conditions were carried out following the manufacturer's recommendations (Invitrogen; no. C11397). Slides were examined with an Olympus BX61 fluorescence microscope and digitally photographed. The images were captured using the Olympus IPP software package and visualized in Photoshop CS 6.0 version.

RNA-Seq Data Processing and Analysis

Total RNAs were isolated from the leaves using Trizol reagent (Invitrogen, United States) according to the manufacturer's instructions. The integrity, quality, and concentration of extracted RNAs were assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies, United States). Transcriptome libraries were constructed for each sample and sequenced using the Illumina HiSeq 2000 platform with standard protocols. Three biological replications were used for the hexaploid plants, and three technical replications were used for the nonaploid plant.

The hexaploid wheat (Chinese Spring) reference genome sequence and its annotation information were downloaded from IWGSC[1]. Each set of cleaned data was aligned to the reference using HISAT2 (version 2.0.1; Kim et al., 2015). The clean data information and mapping rates are shown in **Supplementary Table S1**. The uniquely mapped reads to the reference sequence were computed. The differentially expressed genes (DEGs) were determined by using Cuffdiff (version 2.2.1) by comparing the FPKM values. Transcripts with an FDR-adjusted (Benjamini and Hochberg, 1995) $p < 0.05$ and fold change > 2 were considered to exhibit statistically significant expression differences between samples.

We utilized triad genes that had a 1:1:1 correspondence across the three homoeologous subgenomes in hexaploid common wheat, the definition of homoeolog expression patterns was as reported (Ramirez-Gonzalez et al., 2018). We compared the seven homoeolog expression bias categories in the nonaploid relative to those in the hexaploid wheat.

Gene ontology (GO) enrichment analysis was performed by hypergeometric distribution in R (version 3.4.0) with an adjusted $p < 0.05$ as a cutoff to determine significantly enriched GO terms.

Pyrosequencing

The protocol essentially followed the original report (Mochida et al., 2003) with modifications (Zhang et al., 2013). A set of balanced expressed triads were arbitrarily selected to design the pyrosequencing primers for the purpose of assaying subgenome-specific expression by the pyrosequencing system (PyroMarkID Q96, Qiagen, Germany). The SeqMan program[2] was used to identify subgenome-specific single-nucleotide polymorphisms (SNPs) that enable reliable distinction of the A, B, and D subgenomes for a given triad gene. Consequently, both pyrosequencing primers and gene-specific PCR amplification primers were designed successfully for a set of 18 triads using the Soft Assay Design software. Biotin-labeled PCR products were immobilized on streptavidin-coated paramagnetic beads. Capture of biotinylated single-strand PCR products, annealing of the sequencing primer, and solid-phase pyrosequencing were performed following the manufacturer's recommendations.

RESULTS

The Nonaploid Wheat Contained a Balanced Genome Constitution Yet Manifested Severe Organismal Abnormality in Growth and Development

We used a combinatorial FISH and GISH procedure that allows unequivocal identification of each of the 21 wheat chromosome pairs (Zhang et al., 2013) to study chromosomal stability in various types of newly synthesized wheat. We analyzed a fifth-selfed generation (S_5) plant population that originated from an individual plant of a synthetic allohexaploid wheat (dubbed AT5, $2n = 6x = 42$, BBAADD) formed by crossing a durum wheat (*T. turgidum*, ssp. *durum*, cv. TTR04, $2n = 4x = 28$,

[1]https://www.wheatgenome.org/

[2]http://www.dnastar.com/

BBAA) and *Aegilops tauchii* (accession TQ27, $2n = 2x = 14$, DD), followed by colchicine-induced chromosome doubling. In the course, we identified a spontaneously formed nonaploid plant among 350 karyotyped individuals. This nonaploid plant contained three complete sets of the A, B, and D subgenomes ($2n = 9x = 63$, BBBAAADDD) of hexaploid common wheat with no evidence of numerical or structural chromosomal abnormality across all the 15 root-tip cells examined (**Figures 1A–D**). This suggests the nonaploid plant resulted from a fusion of an unreduced 2n gamete ($6x$) with a normal 1n gamete ($3x$) and with normal mitotic cell divisions. The individual nonaploid seed was not recognized because it appeared normal and showed no difference from those of the hexaploids, which also germinated regularly. However, compared with its isogenic hexaploid siblings, the germinated nonaploid seedling plant manifested conspicuous retardation in growth and development soon after the second-leaf stage with fewer leaves and a single tiller 45 days postgermination (**Figure 1E**). Moreover, it did not show further growth henceforth, failed to enter the reproductive stage, and died around 60 days postgermination.

Massive Global Dysregulation of Gene Expression in Nonaploid Wheat

The impact of genome multiplication *per se* on gene expression in plants remains controversial and tends to be case-specific (Spoelhof et al., 2019). To test whether the abnormal growth and development of the nonaploid wheat were associated with major changes in gene expression due to genome triplication, we conducted a deep RNA-seq-based transcriptome analysis. We used the fully expanded seconf leaf (Simmons et al., 1985) as the target tissue because at this developmental phase no discernible difference between seedlings and the leaves was seen between the nonaploid and its isogenic euploid hexaploid siblings although abnormality in the nonaploid appeared henceforth. Assessing the obtained RNA-seq data indicated that reads of the single nonaploid plant exceeded 109 million while those of the hexaploid plants (with three biological replications) reached up to 188 million. Nevertheless, this difference in sequence depth should not impede reliable comparative analysis because 109 million was already equivalent to 30x coverage of a haploid genome for hexaploid wheat. The reads of each sample were mapped to the updated version of the hexaploid common wheat (cv. CS) reference genome (International Wheat Genome Sequencing Constorium et al., 2018) with an average mapping rate of *ca.* 90%. Correlation coefficients across the three biological replicates of the hexaploid plants were 0.96 ∼ 0.98 (**Supplementary Table S1**), indicating high robustness of our RNA-seq data and analyses.

Collectively, 49,436 genes were found to be expressed in the hexaploid and nonaploid leaf samples (**Table 1**). We used the Cuffdiff software to quantify the DEGs between the nonaploid and hexaploid samples. Surprisingly, we found that, at a twofold change cutoff, 12,454 genes (25.2% of 49,436) were differentially expressed in the nonaploid leaf tissue compared with that of its isogenic hexaploid siblings. Of these DEGs, 7,998 (64.2%),

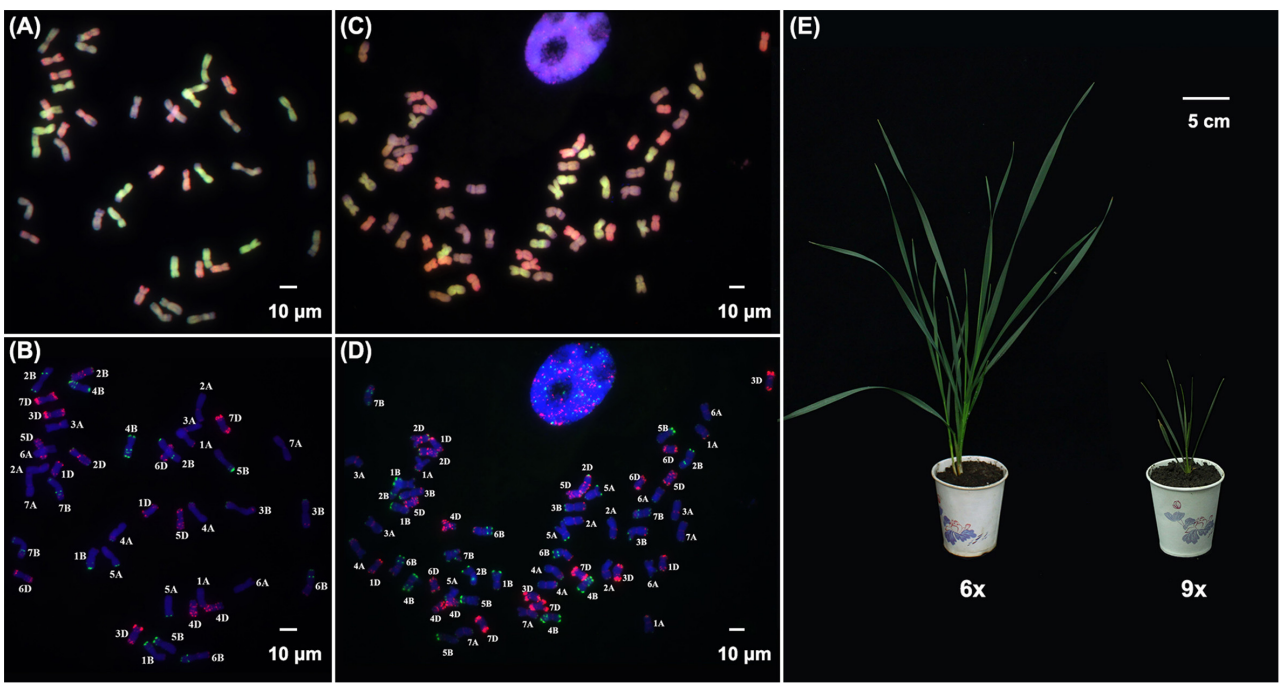


FIGURE 1 | Karyotype and phenotype in hexaploid and nonaploid wheat. **(A)** and **(B)**, FISH and GISH images of a metaphase cell of hexaploid wheat. **(C)** and **(D)**, FISH and GISH images of a metaphase cell of nonaploid wheat. **(E)** Seedlings of hexaploid and nonaploid wheat. The FISH probes used are pSc119.2 (green), and pAS1 (red), the GISH probes used are genomic DNA from *Triticum urartu* (genome AA, green), and *Aegilops tauchii* (genome DD, red); and genome BB (blue) was counterstained with DAPI.

TABLE 1 | The numbers of total expressed genes (EGs) and differentially expressed genes (DEGs) in nonaploid vs. hexaploid wheat plants.

Sample	EGs	DEGs*	Upregulated	Downregulated
9x vs. 6x	49,436	12,454	7,998	4,456

DEG: fold-change > 2.*

and 4,456 (35.8%) were up- and down-regulated, respectively, in the nonaploid (**Table 1** and **Figure 2A**) with the former being significantly more than the later (prop test, $p < 2.2e{-}16$). The DEGs were more or less evenly distributed among the 21 chromosome pairs, indicating that balanced triplication of the wheat genome resulted in large-scale global changes of gene expression (**Figure 2B**).

Notably, among the DEGs, 357 genes were non-expressed in hexaploid but activated in nonaploid, and 100 genes expressed in hexaploid were silenced in nonaploid (**Supplementary Table S3**), further pointing to a largely disruptive effect of genome triplication on transcriptional gene regulation.

To test whether specific cellular or organismal functions might have been impacted by the massive changes of gene expression in general and the nonaploid vs. hexaploid activated or silenced genes in particular, we conducted GO analyses for the various groups of DEGs. We found the up- and down-regulated DEGs, each as a group, showed distinct GO enrichments. The quantitatively upregulated DEGs were mainly enriched in cellular functions related to post-transcriptional regulation, including cytoplasm function, ribosome biogenesis, protein refolding and ubiquitination, and hydrolase activity, and the enriched GO terms for the down-regulated DEGs were mainly involved in organismal growth and development, including photosynthesis, signal transduction, response to hormones (**Figure 2C**). Notably, although the 357 nonaploid vs. hexaploid activated genes did not show significant enrichment, the 100 nonaploid vs. hexaploid silenced genes were overrepresented by GO terms involved in photosynthesis and auxin-response pathways, suggesting the qualitative misregulation of these genes (shutdown) may have played a major part in the impaired growth and development arrest of the nonaploid plant (**Supplementary Figure S1**).

Only Moderate Perturbation of Relative Subgenome Homeologous Expression Patterns in the Nonaploid

It has been established that gene expression of the three homoeologous subgenomes of hexaploid wheat is highly coordinated (Ramirez-Gonzalez et al., 2018). To explore whether balanced triplication of the three subgenomes would disrupt the inherent subgenome expression patterns established in hexaploid wheat, we analyzed the triad genes that had a 1:1:1 correspondence across the three subgenomes in hexaploid wheat (Ramirez-Gonzalez et al., 2018). In the wheat cv. Chinese Spring (CS) reference genome, there were 17,753 sets of expressed triads (i.e., 53,259 genes) in total, including 1,007 non-syntenic triad sets (Ramirez-Gonzalez et al., 2018). We focused on the 9,574 sets of syntenic triads only because we found they were expressed in leaf tissue of both our nonaploid and hexaploid plants. The fragments per kilobase of exon model per million mapped fragments (FPKM) values of these genes were used for cluster analysis. We found that the subgenomes of both nonaploid and hexaploid were preferentially clustered together based on

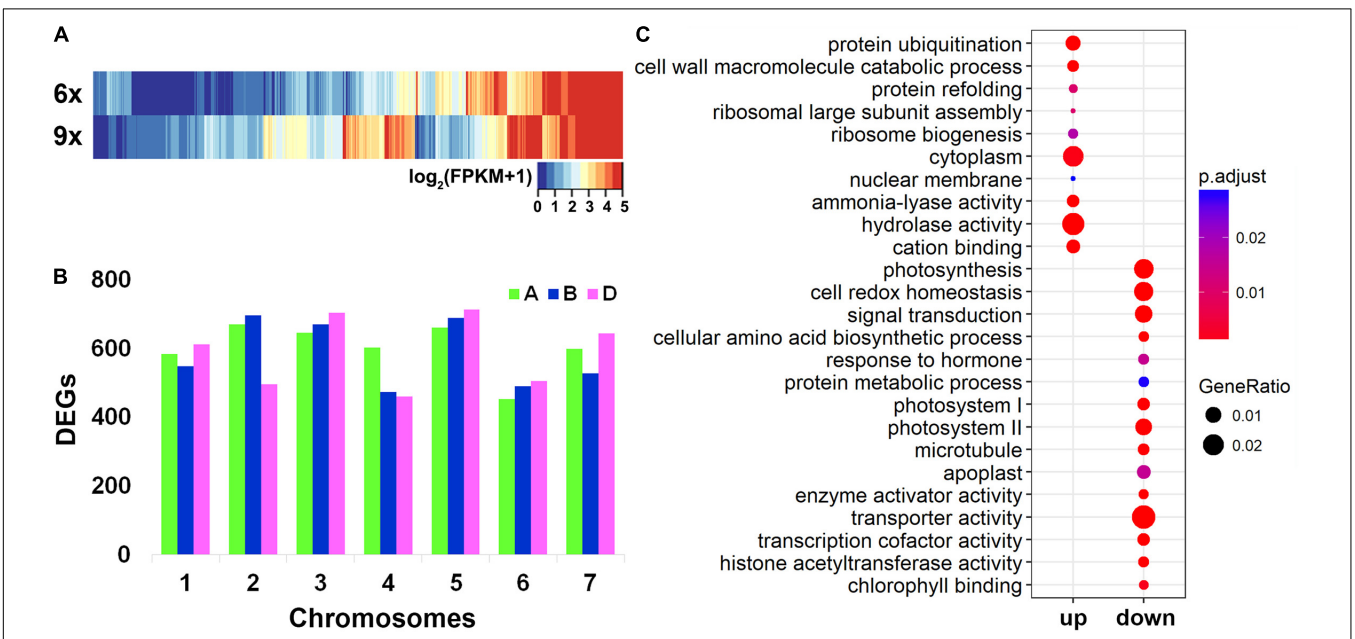


FIGURE 2 | Distribution of differentially expressed genes (DEGs) in chromosomes and gene ontology (GO) analysis. **(A)** A heat map of DEGs according to the FPKM values between hexaploid and nonaploid wheat. **(B)** The distribution of DEGs in the 21 wheat chromosome pairs. **(C)** Enriched GO by up- and downregulated DEGs, respectively.

overall subgenome-specific expression level (relative subgenome transcript abundance) similarities (**Supplementary Figure S2**), suggesting balanced genome triplication did not cause a general disruption of the evolved homoeologous expression patterns (subgenome partitioning) of hexaploid wheat.

According to the prior established classification criteria (Ramirez-Gonzalez et al., 2018), we divided the triad genes into two major categories: balanced, i.e., those with a similar (statistically equal) relative abundance of transcripts from the three subgenome homeologs, and unbalanced, which were further divided into six subcategories, including homeolog dominant or suppressed, depending on the relative higher or lower abundance of transcripts from a single homeolog relative to those of the other two (**Figure 3B** and **Supplementary Figure S3**). Of the total 9,574 sets of expressed triads, the balanced category occupied the predominant proportions in both nonaploid and hexaploid, which were 7,419 (77.5%) and 6,951 (72.6%), respectively; the proportions of the single-homeolog dominant expression pattern were of the lowest category in both nonaploid (4.3%) and hexaploid (6.1%), and proportions of the single-homeolog suppressed category were in between for both nonaploid (18.3%) and hexaploid (21.3%). These proportions of triad expression patterns were also very similar to those seen in the same tissue of CS plants grown under our normal condition (**Figure 3A**). These results indicate that the proportions of triads manifesting each type of homoeologous expression patterns are highly conserved between the nonaploid and hexaploid wheat.

Nevertheless, conservation in proportion does not equate to the conservation of gene content or identity in each of the triad gene expression patterns between the nonaploid and hexaploid. To resolve this issue, we traced the expression pattern for each of the 7,775 sets of triads from hexaploid to nonaploid. We found homoeologous expression patterns in 81.2% (7,775 out of 9,574) of the triads remained the same between hexaploid and nonaploid while those in the remaining 18.8% (1,799 out of 9,574) triads changed to different patterns (**Figure 3C**). Of note, the pattern-changed triads mainly concerned the suppressed and dominant categories, which shifted to the balanced category or vice versa with few triads manifesting shifting to direct opposite patterns, i.e., from suppressed to dominant or vice versa (**Figure 3C**).

Taken together, it seemed that balanced genome triplication in wheat did not invoke a major perturbation to the highly coordinated subgenomic or homeologous expression patterns in hexaploid wheat (Ramirez-Gonzalez et al., 2018). To experimentally test the reliability of our subgenome expression patterns as well as of our RNA-seq analyses in general, we performed locus-specific cDNA pyrosequencing for a subset of 18 triads that belonged to the balanced category, which harbored diagnostic SNPs that enabled the design of pyrosequencing primers (**Supplementary Table S2**). For each of the 18 triads, the

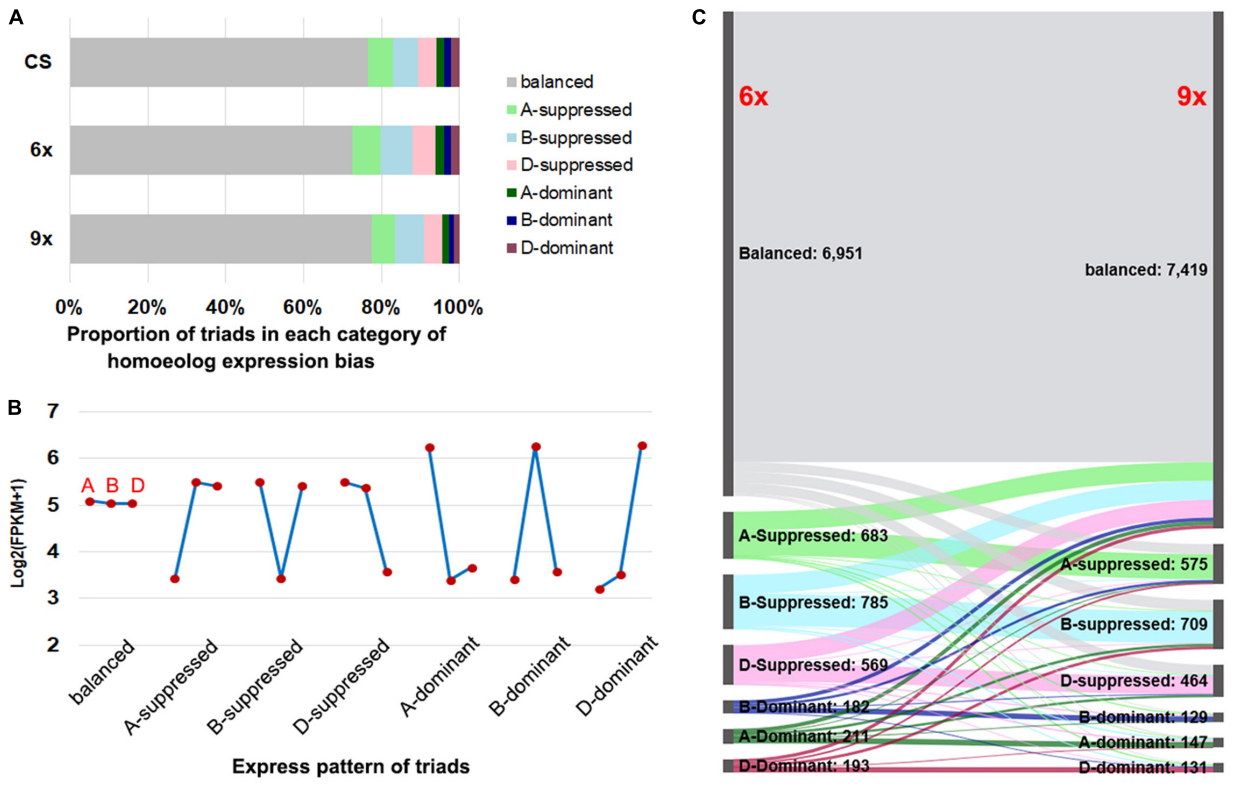


FIGURE 3 | The expression pattern of triad genes between hexaploid and nonaploid wheat. **(A)** Proportion of triads in seven categories of homeolog expression bias in CS, hexaploid and nonaploid plants. **(B)** The model of average FPKM values for each subgenome from seven categories. **(C)** Alluvial plot of classification of triads between hexaploid and nonaploid wheat.

ratio of transcripts from the three subgenomes was determined. We found that in all 18 triads the pyrosequencing data were in line with the RNA-seq data (**Supplementary Figure S4**).

DISCUSSION

In this study, we compared a spontaneously formed nonaploid wheat individual with its isogenic hexaploid siblings at multiple levels, karyotype, phenotype, and gene expression. We found that the nonaploid individual maintained a stable somatic karyotype with no numerical or structural chromosomal variations detected in any of the 15 well-spread root-tip metaphase cells examined, suggesting mitosis was not impaired in the nonaploid with a balanced genome triplication. However, the nonaploid showed severe growth retardation and overall morphological abnormality immediately after the second-leaf stage, leading to premature death. In parallel, large-scale alteration of transcriptional gene expression was detected in the nonaploid relative to its isogenic hexaploid siblings in leaf tissue before the manifestation of phenotypic abnormality, suggesting gene dysregulation precedes the phenotypic abnormality. Due to limitation of material from the single nonaploid plant, we could not conduct experiments assessing possible differences in transcriptome sizes (Doyle and Coate, 2019) between the isogenic nonaploid and hexaploid plants. However, according to recent studies (Spoelhof et al., 2019; Visger et al., 2019), transcriptome size can be significantly altered due to an increase in ploidal level, and when this factor was taken into account, a substantially greater number of DEGs were identified (Visger et al., 2019). Thus, it is reasonable to deduce that the number of DEGs we identified in the nonaploid vs. hexaploid comparisons is an underestimate.

Notably, albeit there was upheaval of the overall gene expression level in the nonaploid, the tightly controlled inter-subgenome relative expression in hexaploid wheat (Ramirez-Gonzalez et al., 2018) was largely maintained. This is in contrast with the situation of aneuploidy in hexaploid wheat in which numerical change of a single chromosome caused large perturbation of subgenome expression (Zhang et al., 2017), suggesting coordinated subgenome regulation is determined by genome balance rather than genome dosage (Birchler and Veitia, 2012). This upheaval of overall gene expression vs. largely stable subgenome relative expression is also consistent with the possibility that changes in epigenetic regulation have played a role. Chromatin epigenetic modifications, including DNA methylation, histone modification and differential titration of regulatory small RNAs, are known to undergo variations due to polyploidization (Song and Chen, 2015). Given that epigenetic modifications are brought about by specific enzymatic machinery that acts *in trans* (Springer et al., 2016), it is conceivable that, in cases of allopolyploidy, they would affect all constituent subgenomes and, hence, cause alterations in overall gene expression rather than in subgenome-specific or -preferred expression.

Due to severe retardation of growth/development and inviability of the single nonaploid plant, we were also unable to perform more intricate analyses, such as measuring cell size/number, cell division rate, or cell wall composition as was done in the *A. thaliana* ploidal series (Corneillie et al., 2019). However, it is conceivable that perturbation of similar physiological and cell biological attributes typical of high ploidy syndrome (Tsukaya, 2008) might have been incurred in the nonaploid wheat although, in our case, the syndrome is apparently more drastic as it culminates in premature lethality. We can rule out disrupted mitosis as a cause for the syndrome (Comai, 2005) as all examined metaphase cells are euploidy, and also, no lagging chromosome was detected at anaphase. Thus, the ploidy syndrome in our case is most probably due to disruption of overall gene regulation, which affected a substantial proportion of the expressed genes, including activation and silencing of critical genes involved in both fundamental cellular activity and tissue/organ growth at the organismal level. It can be envisioned that the dysregulation of gene expression is likely to be further exacerbated following the manifestation of the ploidy syndrome due to component scaling disproportionality and geometrical stress (Storchova et al., 2006). This vicious cycle of dysregulated gene expression and anomalous physiology/cellular structure might have led to inviability of the nonaploid plant.

It is intuitive that each species has an upper limit of ploidal level, across which the high ploidy syndrome will appear, and the severity thereof will scale with further increase of ploidal levels, but the exact ploidal thresholds may vary markedly in different species. It is also conceivable that manifestation of high ploidy syndrome depends on genome constitution of the polyploid in question. For example, octoploid *Triticale* (BBAADDRR, $2n = 8x = 56$) can be readily created by crossing hexaploid wheat with diploid rye (*Secale cereale* L.) followed by chromosome doubling, and which is vigorous and fully fertile. Another relevant issue is whether the three subgenomes of hexaploid wheat are equally sensitive to multiplication. A previous study showed that induced autotetraploid of *Triticum monococcum* (A^mA^m) is overall smaller than their diploid counterparts although the tetraploid plants are highly fertile with normal mitosis and even show near regular meiosis (Kuspira et al., 1985). This suggests at least the A subgenome of hexaploid wheat, which is derived from *T. urartu* and is highly similar to A^m of *T. monococcum*, is likely sensitive to triplication due to reasons we report here. Curiously, however, we did not find that genes located to the A subgenome are particularly disturbed (as DEGs), suggesting the B and D subgenomes are probably equally sensitive to triplication. In conclusion, this study shows that triplication of the three subgenomes of hexaploid wheat causes massive disruption of overall gene expression and imposes a severely detrimental effect on growth and development at the vegetative stages. Our results, thus, suggest hexaploidy is likely the upper limit of ploidal level in wheat. We should caution, however, that the occurrence and magnitude of abnormality associated with genome triplication in wheat may also depend on genetic backgrounds and, thus, different genotypes may vary. This will remain an open question until additional nonaploid or higher ploidal level plants can be obtained in different hexaploid wheat genotypes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in NCBI SRA accession PRJNA624803.

AUTHOR CONTRIBUTIONS

XG and BL: designed the study. XG, RL, CW, TF, and YS: analyzed the data and performed the experiments. XG, LG, HZ, and BL: summarized the data and wrote the manuscript. All authors contributed to this study, read and approved the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (NSFC #31830006) and Natural Science Foundation of Jiangsu Province (Grant/Award #BK20190996), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00687/full#supplementary-material

FIGURE S1 | Gene Ontology of nonaploid vs. hexaploid activated and silenced genes. **(A)** Cellular component. **(B)** Biological process.

FIGURE S2 | Heat map of expressed syntenic triads according the FPKM values between hexaploid and nonaploid wheat. Subgenomes of both nonaploid and hexaploid were preferentially clustered together. The color key is indicated at the bottom.

FIGURE S3 | The relative contribution to total gene expression by each subgenome based on triad assignment to the seven categories (detailed in Main text) between hexaploid and nonaploid wheat.

FIGURE S4 | Analysis of subgenome expression partitioning by locus-specific cDNA pyrosequencing of 18 triads belonging to the balanced category and its comparison with analysis using the RNA-seq data.

TABLE S1 | Details of the clean data generated in the RNA-seq data set.

TABLE S2 | The list of triads used in pyrosequencing.

TABLE S3 | A list of activated and silenced genes in nonaploid vs. hexaploid.

REFERENCES

Adams, K. L. (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. *J. Hered.* 98, 136–141. doi: 10.1093/jhered/esl061

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Birchler, J. A., and Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14746–14753. doi: 10.1073/pnas.1207726109

Blischak, P. D., Mabry, M. E., Conant, G. C., and Pires, J. C. (2018). Integrating networks, phylogenomics, and population genomics for the study of polyploidy. *Annu. Rev. Ecol. Evol. Syst.* 49, 253–278. doi: 10.1146/annurev-ecolsys-121415-032302

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Corneillie, S., De Storme, N., Van Acker, R., Fangel, J. U., De Bruyne, M., De Rycke, R., et al. (2019). Polyploidy affects plant growth and alters cell wall composition. *Plant Physiol.* 179, 74–87. doi: 10.1104/pp.18.00967

Doyle, J. J., and Coate, J. E. (2019). Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* 180, 1–52. doi: 10.1086/700636

Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316, 1862–1866. doi: 10.1126/science.1143986

Feldman, M., and Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat. *Genetics* 192, 763–774. doi: 10.1534/genetics.112.146316

Feldman, M., and Levy, A. A. (2015). "Origin and evolution of wheat and related triticeae species," in *Alien Introgression in Wheat*, eds M. Molnár-Láng, C. Ceoloni, and J. Doležel (Cham: Springer).

Gjelsvik, K. J., Besen-McNally, R., and Losick, V. P. (2019). Solving the polyploid mystery in health and disease. *Trends Genet.* 35, 6–14. doi: 10.1016/j.tig.2018.10.005

International Wheat Genome Sequencing Constorium, Investigators, I. R. P., Appels, R., Eversole, K., Feuillet, C., Keller, B., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Jiao, Y., and Paterson, A. H. (2014). Polyploidy-associated genome modifications during land plant evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130355. doi: 10.1098/rstb.2013.0355

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Kashkush, K., Feldman, M., and Levy, A. A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160, 1651–1659. doi: 10.0000/PMID11973318

Kato, A., Lamb, J. C., and Birchler, J. A. (2004). Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13554–13559. doi: 10.1073/pnas.0403659101

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kuspira, J., Bhambmani, R. N., and Shimada, T. (1985). Genetic and cytogenetic analyses of the A genome of *Triticum monococcum*. I. Cytology, breeding behaviour, fertility, and morphology of induced autotetraploids. *Can. J. Genet. Cytol.* 27, 51–63. doi: 10.1139/g85-010

Madlung, A., and Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet. Genome Res.* 140, 270–285. doi: 10.1159/000351430

Marques, D. A., Meier, J. I., and Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34, 531–544. doi: 10.1016/j.tree.2019.02.008

Matsuoka, Y. (2011). Evolution of polyploid triticum wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol.* 52, 750–764. doi: 10.1093/pcp/pcr018

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

Mochida, K., Yamazaki, Y., and Ogihara, Y. (2003). Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genomics* 270, 371–377. doi: 10.1007/s00438-003-0939-7

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798

Ramirez-Gonzalez, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:eaar6089. doi: 10.1126/science.aar6089

Robinson, D. O., Coate, J. E., Singh, A., Hong, L., Bush, M., Doyle, J. J., et al. (2018). Ploidy and size at multiple scales in the *Arabidopsis* Sepal. *Plant Cell* 30, 2308–2329. doi: 10.1105/tpc.18.00344

Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R., and Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* 3, 429–441. doi: 10.1038/nrg817

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15:62. doi: 10.1186/s12915-017-0399-x

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

Simmons, S. R., Oelke, E. A., and Anderson, P. M. (1985). *Growth and Development Guide for Spring Wheat*. St. Paul: University of Minnesota Agricultural Extension Service.

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Soltis, P. S., and Soltis, D. E. (2012). *Polyploidy and Genome Evolution*. Cham: Springer.

Song, Q., and Chen, Z. J. (2015). Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* 24, 101–109. doi: 10.1016/j.pbi.2015.02.007

Spoelhof, J. P., Keeffe, R., and McDaniel, S. F. (2019). Does reproductive assurance explain the incidence of polyploidy in plants and animals?. *New Phytol.* 227, 14–21. doi: 10.1111/nph.16396

Springer, N. M., Lisch, D., and Li, Q. (2016). Creating order from chaos: epigenome dynamics in plants with complex genomes. *Plant Cell* 28, 314–325. doi: 10.1105/tpc.15.00911

Storchova, Z., Breneman, A., Cande, J., Dunn, J., Burbank, K., O'Toole, E., et al. (2006). Genome-wide genetic analysis of polyploidy in yeast. *Nature* 443, 541–547. doi: 10.1038/nature05178

Tsukaya, H. (2008). Controlling size in multicellular organs: focus on the leaf. *PLoS Biol.* 6:e174. doi: 10.1371/journal.pbio.0060174

Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L., and Vandepoele, K. (2009). The flowering world: a tale of duplications. *Trends Plant Sci.* 14, 680–688. doi: 10.1016/j.tplants.2009.09.001

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

Visger, C. J., Wong, G. K., Zhang, Y., Soltis, P. S., and Soltis, D. E. (2019). Divergent gene expression levels between diploid and autotetraploid Tolmiea relative to the total transcriptome, the cell, and biomass. *Am. J. Bot.* 106, 280–291. doi: 10.1002/ajb2.1239

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Yoo, M. J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Zhang, A., Li, N., Gong, L., Gou, X., Wang, B., Deng, X., et al. (2017). Global analysis of gene expression in response to whole-chromosome aneuploidy in hexaploid wheat. *Plant Physiol.* 175, 828–847. doi: 10.1104/pp.17.00819

Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., et al. (2013). Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3447–3452. doi: 10.1073/pnas.1300153110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 28 July 2020
doi: 10.3389/fgene.2020.00792

Genomics of Evolutionary Novelty in Hybrids and Polyploids

Gonzalo Nieto Feliner[1], Josep Casacuberta[2] and Jonathan F. Wendel[3]*

[1] Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Madrid, Spain, [2] Center for Research in Agricultural Genomics, CRAG (CSIC-IRTA-UAB-UB), Barcelona, Spain, [3] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, United States

It has long been recognized that hybridization and polyploidy are prominent processes in plant evolution. Although classically recognized as significant in speciation and adaptation, recognition of the importance of interspecific gene flow has dramatically increased during the genomics era, concomitant with an unending flood of empirical examples, with or without genome doubling. Interspecific gene flow is thus increasingly thought to lead to evolutionary innovation and diversification, via adaptive introgression, homoploid hybrid speciation and allopolyploid speciation. Less well understood, however, are the suite of genetic and genomic mechanisms set in motion by the merger of differentiated genomes, and the temporal scale over which recombinational complexity mediated by gene flow might be expressed and exposed to natural selection. We focus on these issues here, considering the types of molecular genetic and genomic processes that might be set in motion by the saltational event of genome merger between two diverged species, either with or without genome doubling, and how these various processes can contribute to novel phenotypes. Genetic mechanisms include the infusion of new alleles and the genesis of novel structural variation including translocations and inversions, homoeologous exchanges, transposable element mobilization and novel insertional effects, presence-absence variation and copy number variation. Polyploidy generates massive transcriptomic and regulatory alteration, presumably set in motion by disrupted stoichiometries of regulatory factors, small RNAs and other genome interactions that cascade from single-gene expression change up through entire networks of transformed regulatory modules. We highlight both these novel combinatorial possibilities and the range of temporal scales over which such complexity might be generated, and thus exposed to natural selection and drift.

Keywords: adaptation, allopolyploidy, gene and genome duplication, transposable elements, hybridization, phenotypic novelty, radiation lag-time model

OPEN ACCESS

Edited by:
Jianping Wang,
University of Florida, United States

Reviewed by:
Jeffrey Joseph Doyle,
Cornell University, United States
Douglas Soltis,
University of Florida, United States
Jeremy Coate,
Reed College, United States, in collaboration with reviewer JD

****Correspondence:***
Gonzalo Nieto Feliner
nieto@rjb.csic.es

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *28 April 2020*
Accepted: *03 July 2020*
Published: *28 July 2020*

Citation:
Nieto Feliner G, Casacuberta J and Wendel JF (2020) Genomics of Evolutionary Novelty in Hybrids and Polyploids. Front. Genet. 11:792. doi: 10.3389/fgene.2020.00792

INTRODUCTION

One of the remarkable realizations of the genomics era is that hybridization—crosses between individuals from populations that are distinguishable on the basis of one or more heritable characters (Harrison, 1990)—and interspecific gene flow—the successful movement of genes among populations (Ellstrand, 2014)—are far more prevalent than previously recognized.

Genomic data from a wide range of model and non-model organisms have both confirmed and greatly extended the pre-existing notion that natural hybridization is a frequent phenomenon in the living world, and not only in plants (Mallet, 2005; Soltis and Soltis, 2009; Green et al., 2010; Fontaine et al., 2015; Leducq et al., 2016; Elgvin et al., 2017; Meier et al., 2017; Lamichhaney et al., 2018), and that it often leads to interspecific gene flow. The observation that hybridization is associated with allopolyploid speciation has long been recognized, and in fact this became fundamental in plant evolutionary thinking, with the original evidence primarily consisting of cytological observations that peaked in the 1970s (Stebbins, 1940, 1971; Grant, 1981). The advent of molecular markers magnified this interest, stimulating research into polyploidy across the phylogenetic spectrum of angiosperms and using both natural and synthetic allopolyploids (reviewed in, e.g., Soltis and Soltis, 2012; Soltis et al., 2015, 2016; Van de Peer et al., 2017). This interest continues apace today, as evidenced by the present special issue of *Frontiers*.

Among the more surprising insights to emerge from this research is that the genomes of *all* modern angiosperms contain vestiges of multiple past rounds of polyploidy (Jiao et al., 2011; Soltis et al., 2016; Leebens-Mack et al., 2019; but see Ruprecht et al., 2017a), some ancient and in many cases some quite recent, with each event superimposed on the genomic remnants of earlier rounds of doubling. A second realization is that each whole genome doubling (WGD) event has been followed by incompletely understood genome fractionation processes (Wendel, 2015; Soltis et al., 2016; Bird et al., 2018; Cheng et al., 2018; Wendel et al., 2018), as well as myriad immediate and longer-term responses to genome merger and doubling at the genomic, expression, and cellular levels (Yoo et al., 2014; Wendel, 2015; Soltis et al., 2016; Sharbrough et al., 2017; Doyle and Coate, 2019). Thus, the architecture of modern plant genomes reflects, in part, the residuals from the superimposed joint action of WGD and fractionation, these twin processes encompassing the "wondrous cycles of polyploidy" (Wendel, 2015).

One consequence of the pervasiveness of polyploidy and its dynamism over time is the widely held view that whole-genome doubling plays an important role in generating phenotypic novelty. This topic has been the subject of speculation for decades (Levin, 1983; Soltis, 2013; Soltis P. S. et al., 2014; Vanneste et al., 2014; Edger et al., 2015; Tank et al., 2015; Van de Peer et al., 2017), but to date the number of cases where polyploidy itself has been convincingly connected to specific phenotypic innovations remains relatively small. Part of the challenge in demonstrating this connection is that adaptation and diversification take place over a diverse spectrum of time-scales, as do the various genomic diversification and fractionation processes set in motion by polyploidy.

In addition to hybridization *with* WGD (allopolyploidy), evidence abounds for the occurrence of homoploid hybridization, that is, hybridization and gene flow *without* WGD (Mallet, 2007; Pennisi, 2016; Runemark et al., 2019). Three decades of molecular phylogenetic studies (e.g., Sang et al., 1995; Barrier et al., 1999; Blanco-Pastor et al., 2012; García et al., 2017; Marques et al., 2017; Folk et al., 2018) and more recent genomic scrutiny (Baack and Rieseberg, 2007; Twyford and Ennos, 2012; Abbott et al., 2016; Payseur and Rieseberg, 2016) have revealed numerous hybrid lineages across the living world. Ever since hybridization and introgression started to be inferred from incongruence between gene-trees, it has been realized that an alternative neutral process—incomplete lineage sorting (ILS) or deep coalescence—could lead to similar phylogenetic patterns. Favored by large populations sizes and short speciation times, the occurrence of ILS is still often inferred whenever evidence for introgression is lacking. However, a number of methodological approaches, the most widely used of which is the ABBA-BABA test, are now available to tease apart the two phenomena (Joly et al., 2009; Green et al., 2010; Blanco-Pastor et al., 2012). What remains a matter of some debate is how often hybridization plays a creative role in evolution, or phrased alternatively, whether the advantages accrued from natural hybridization are responsible for the pervasiveness of this phenomenon. The influential ideas of Mayr (1963) that hybridization was an evolutionary dead-end began to be challenged after breakthrough discoveries in birds and insects (Mavárez and Linares, 2008; Grant and Grant, 2009, 2017; Salazar et al., 2010), but the relative importance of hybridization in lineage diversification remains a subject of active debate. Some posit that hybridization, even if frequent, is likely transient in genomes and thus of little evolutionary relevance (Barton, 2013; Servedio et al., 2013); according to this view, merged genomes mostly evolve in the direction of purging incompatibilities (Schumer et al., 2016, 2018). The opposite view, that natural hybridization may contribute positively to adaptation, differentiation and speciation, is embraced by many empirically oriented evolutionary biologists (Rieseberg, 1991; Arnold, 1993; Wang et al., 2001; Mallet, 2007; Abbott et al., 2010, 2013; Butlin and Ritchie, 2013; Sætre, 2013; Soltis, 2013; Yakimowski and Rieseberg, 2014; Grant and Grant, 2017; Nieto Feliner et al., 2017; Ottenburghs, 2018; Wagner, 2018), impressed as they are by the ever-increasing number of discoveries of gene-tree conflict in datasets. In part, these two views are fueled by the contrast between the burgeoning number of lineages that are unveiled by molecular phylogenetic studies to have a hybrid ancestry and the tiny fraction of cases in which we understand how hybridization may have led to adaptation (and/or phenotypically relevant drift) and diversification (Schumer et al., 2014). We suspect that this scarcity of well-understood examples reflects both insufficiency in our understanding of the genetic bases of adaptive traits and their inherent diversity and context dependency, as well as the temporal disconnect between hybridization events (ancient and recent) and adaptation to ecological conditions that may no longer be present.

Allopolyploidization and homoploid hybrid speciation, of course, comprise just two of the many possible evolutionary outcomes of interspecific genetic exchange (Runemark et al., 2019). Others include reinforcement (Hopkins, 2013), genetic assimilation (Levin et al., 1996; Ehrenreich and Pfennig, 2016), formation of various kinds of hybrid zones (Barton and Hewitt, 1985; Harrison, 1993; Abbott, 2017), and adaptive or neutral introgression (Rieseberg and Wendel, 1993; Heliconius Genome Consortium, 2012; Schmickl et al., 2017; Suarez-Gonzalez et al., 2018b; Edelman et al., 2019). These many possibilities serve to

illustrate both the prevalence and complexity of the outcomes of secondary contact among divergent lineages, and additional, seemingly unlikely possibilities continue to emerge. For instance, inter-ploidy gene flow between lineages that "should be" reproductively isolated turns out to characterize the evolution of many allopolyploid lineages or diploid-allopolyploid complexes (Grant, 1981; Zohren et al., 2016; Hohmann and Koch, 2017; Marburger et al., 2019; Monnahan et al., 2019).

There are also connections between hybrid zones and homoploid hybrid speciation (Hodges et al., 1996), and between adaptive introgression and homoploid hybrid speciation (Brower, 2013). In fact, some authors propose that these outcomes of hybridization represent different stages of a continuum of speciation (Seehausen et al., 2014; Lowry and Gould, 2016; Roux C. et al., 2016). This notion illustrates how hybridization usually occurs in a complex spatial and temporal context (Abbott et al., 2013; Sætre, 2013). Multiple different factors, such as levels of divergence and ecological opportunity, may determine whether raw genetic variance introduced by hybridization—two to three orders of magnitude greater than that introduced by mutation, according to Grant and Grant (1994)—facilitates adaptive divergence or contributes to evolutionary novelty. That the ultimate outcomes of natural hybridization and allopolyploidy depend on numerous interacting factors sieved by selection over various timescales and ecological contexts makes predictions extremely difficult (Butlin and Ritchie, 2013), and favors expectations that somehow incorporate stochasticity, e.g., an evolutionary novelty hybrid zone model (Arnold, 1997). Examples of this stochasticity in the short term includes hybrid unviability of some genotypes even in F_1 hybrids between conspecific genotypes of an inbreeding species (Bomblies and Weigel, 2007) and post-F_1 hybrids between species due to Bateson-Dobzhansky-Muller (BDM) incompatibilities and/or breakdown of coadaptive gene complexes following recombination (Christe et al., 2016).

Given the numerous avenues by which genetic exchanges may occur between differentiated genomes, we thought it timely to consider the question of how this merger might ultimately lead to novel phenotypes and adaptation. We leave aside the classic but still debated topic of heterosis, which refers to the process by which hybrids, including allopolyploids, may exhibit greater biomass, speed of development, and fertility than both parents (Birchler et al., 2010; Hochholdinger and Baldauf, 2018). At the outset, hybridization leads to infusion of new genetic material, which is either rapidly removed by selection and/or drift, or partially removed, leaving behind a transformed genome. But in most cases, even involving iconic examples such as *Iris* (Martin et al., 2006), *Helianthus* (Rieseberg et al., 2003), and *Populus* (Suarez-Gonzalez et al., 2016, 2018b) the specific identity and connections between introgressed material and adaptive phenotypes remains elusive or at least incompletely defined. In this light, and perhaps as a form of foreshadowing, we therefore consider the spectrum of novel molecular genetic and genomic processes that may be set in motion by the saltational genomic shock of merger, both with and without genome doubling. Our focus is on the possible selective advantages of hybridized genomes that lead to adaptation, phenotypic novelty, and diversification, as opposed to simply transient effects. Using examples mostly from plants we highlight genetic and genomic consequences of genome merger that create potentially adaptive phenotypes. Our intention is not to be comprehensive nor encyclopedic, but instead to present an overview of the possible mechanisms by which new phenotypes may arise as a result of interspecific gene flow. As an organizational framework, we arrange these effects into three categories, noting that these are not mutually exclusive and often occur in concert: (1) responses at the genetic and genomic level, such as structural diversity or copy number variation; (2) responses at the gene expression and regulatory level, such as neo- and subfunctionalization of duplicated loci; and (3) responses at combinatorial genomic and expression-levels, such as cytonuclear interactions or transposable element activity.

RESPONSES TO GENOME MERGER AT THE GENETIC AND GENOMIC LEVEL

Genic Introgression

Plant genomes vary enormously in virtually every feature used to describe their composition or "suite of residents," including, from the smallest scale to the largest, nucleotide composition, gene and regulatory sequences, genic content and copy numbers, repetitive sequences and transposable element content, chromosome numbers, genome size (Wendel et al., 2016) and a spectrum of epigenetic features. As populations and species diverge, so will their genomes, though not necessarily monotonically across categories of genomic change nor homogeneously among lineages. Nonetheless, the divergence of lineages is inevitably associated with the accumulation of multiple forms of mutational differences, again, small to large. At the simplest level, divergence is associated with changes in gene sequence, either in coding or regulatory sequences, and thus genomic reunions associated with hybridization and/or WGD may lead to new genic contexts with possible effects on selectively relevant phenotypes (**Figure 1**).

In principle, hybridization-induced infusion of diverged or novel genes might be considered the most straightforward form of evolutionarily relevant introgression to detect following hybridization, as it is a straightforward matter now to assay gene sequences. Accordingly, it is not surprising to find recent examples of adaptation or at least phenotypic novelty caused by genic introgression. An elegant early example was from *Senecio* (Kim et al., 2008), who showed that introgression of the RAY locus from the diploid hybrid species *S. squalidus* into the tetraploid *S. vulgaris* causes the formation of bilaterally symmetrical flowers (ray florets) on the periphery of the heads that otherwise are discoid. The introgressed RAY locus comprises a cluster of cycloidea-like genes, which encode DNA-binding proteins known to cause asymmetry in *Antirrhinum* (Luo et al., 1996). One of the more remarkable features of the introgressed RAY locus story is that the gene flow surmounted a seemingly improbable reproductive barrier, from diploid to tetraploid. How this occurs is unclear, but interploidal gene flow is a recurring theme (see below). An added notable

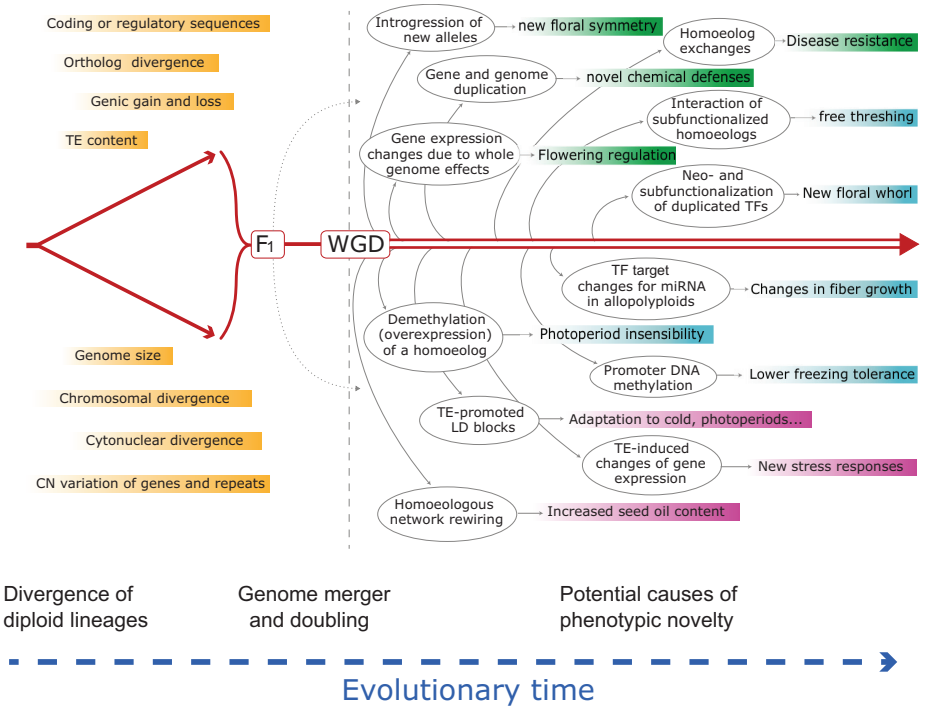


FIGURE 1 | Schematic evolutionary diagram depicting changes accompanying divergence of two diploid lineages, followed by phenomena enabled by or set in motion by hybridization and genome doubling, either separately or together. Shown is a sampling of genetic and genomic mechanisms subsequent to hybridization and an allopolyploid event that could, over time (arrow at bottom), lead to novel adaptation and phenotypes. Many of these phenomena and consequences may be latent for hundreds to millions of years. Box colors indicate categories of change in the order discussed in the text: responses at the genomic and genetic level (green); responses at the gene expression and regulatory level (cyan); responses at the combinatorial genomic and expression levels (mauve). See text for elaboration of these and many other examples. CN, copy number; HE, homoeologous exchange; LD, linkage disequilibrium; miRNA, microRNA; TE, transposable element; TF, transcription factor; WGD, whole genome duplication. At present the relative importance and prevalence of these phenomena following hybridization vs. polyploidy is unknown, but in many cases the two organismal-level processes independently set in motion at least some of the illustrated genomic responses.

dimension to this example is that the gene flow apparently is adaptive, as ray florets promote outcrossing and thereby infuse genetic variation into the otherwise selfing *S. vulgaris* (Kim et al., 2008).

A second beautiful example of inter-ploidal gene flow and adaptation to the polyploid condition is the paper (Marburger et al., 2019) on diploid and autopolyploid *Arabidopsis arenosa* and *A. lyrata* in Europe. Both species have diploid and autopolyploid populations and experience at least occasional diploid-tetraploid gene flow. By resequencing 92 individuals from 30 populations collected from a broad range of both species, Marburger et al. (2019) demonstrate that interspecific introgression has occurred bidirectionally, and that some *A. arenosa* introgression peaks into *A. lyrata* are both narrow and associated with strong signatures of selection. Remarkably, these small regions of interspecific introgression include key genes known to be important in stabilizing meiosis following WGD, suggesting that adaptation to polyploidy was mediated by interspecific gene flow. A fascinating twist on this story is that the *A. arenosa* alleles introgressed into *A. lyrata* are posited to have been favored because WGD in the former species is older than in the latter. Thus, its alleles at meiosis-stabilizing genes are better adapted to the tetraploid condition than the more naïve and native *A. lyrata* genes, and so selection has favored their replacement.

The preceding examples of adaptive genic introgression represent the unusual cases where adaptation has been at least arguably causally connected to specific genes. Far more numerous are examples where genomic evidence for adaptive introgression is convincing, but either the responsible genes have not been identified or they have not been directly linked to phenotypes that are unequivocally connected to adaptation (Suarez-Gonzalez et al., 2016, 2018a; Hübner et al., 2019; Janzen et al., 2019; Mitchell et al., 2019). A recent case in point is for cultivated sunflowers (Hübner et al., 2019), where genomic resequencing of about 400 cultivated lines, Native American landraces and wild accessions from 11 wild species demonstrated that 1.5% of the genes in cultivated sunflower arose via interspecific introgression from wild species. Many of these genes are connected to biotic resistance such as downy mildew resistance, implicating selection for disease resistance as being responsible for adaptive interspecific gene flow. Similar examples of either intentional or unintentional adaptive introgression are common in our major cultivated crops (Janzen et al., 2019).

The preceding examples all involve either introgression between closely related congeners or wild-domesticated

comparisons. In these instances the temporal window for detecting adaptive introgression may be maximized relative to other scales of divergence, in that as the latter increases, along with time since divergence, there is less certainty with respect to the relevant ecology (and hence insight into selective pressure). An interesting recent example in this respect is from Tibetan *Cupressus* (Ma et al., 2019), where a variety of transcriptomic and population genetic tools were used to demonstrate that adaptation to colder and drier environments in one species was enabled via introgression from a second species, perhaps 200,000 years ago. Detecting the footprints of selection will be more difficult, however, as equilibrium is restored following selective sweeps.

Challenges in detecting adaptive introgression are not just restricted to older events, because even recent adaptive introgression may be difficult to distinguish from other causes of patterns of variation. Clinal variation across environmental gradients, for example, may arise from local differentiation, or, as in the case of *Cupressus* cited above (Ma et al., 2019), from asymmetric interspecific introgression (see also, e.g., Welch and Rieseberg, 2002; Rieseberg et al., 2003; Arnold et al., 2010; Scascitelli et al., 2010; Whitney et al., 2010; Leroy et al., 2020). A powerful approach for distinguishing adaptive introgression from other sources of variation entails the combined use of genome-wide tools, now accessible for most non-model organisms, with demographic and population genetic modeling (Pease et al., 2016; Aeschbacher et al., 2017; Martin and Jiggins, 2017; Ma et al., 2019). Genome-wide approaches also provide insight into key questions such as whether introgression is scattered or localized across the genome, how it has been shaped by selection (Suarez-Gonzalez et al., 2016), and whether adaptively introgressed alleles had diverged in the donor species (Parchman et al., 2013; Bay and Ruegg, 2017; Leroy et al., 2020). These considerations are finding increasing utility in the rescue of genetically impoverished or threatened species (Hamilton and Miller, 2016).

These many challenges associated with the passage of time, historical ecological inference, and interpretation of patterns of genetic diversity also apply to our understanding of adaptation following ancient episodes of polyploidy. In most cases, it is unknown whether older WGD events involved autopolyploidy or allopolyploidy, and thus even though it might be clear that there are functionally divergent homoeologs (including subfunctionalized and neofunctionalized), it is not at all evident whether this divergence represents evolution at the polyploid level or whether this reflects merger of pre-existing differences. Comparative genomics often, however, yields important clues, as in the example of the butterfly (Pieridae)-glucosinolate "arms race" in the Brassicales (Edger et al., 2015), where gene and genome duplication is implicated in novel chemical defenses in the plants as well as countermeasures in the butterflies. Similarly suggestive evidence is common in other plant groups (Sato et al., 2012; Vanneste et al., 2014; Lohaus and Van de Peer, 2016), and certain classes of genes and transcription factors involved in stress responses have been found to be repeatedly preferentially retained following WGDs in 25 different angiosperm lineages (Wu et al., 2019). These examples collectively provide tantalizing evidence which implicate, but do not prove, that WGD was responsible for evolutionary specializations or adaptations.

Structural Diversity

Merger of differentiated genomes leads not only to transfer of genes but also to the incorporation of structural variants in a novel genomic context. Structural variants (SVs) include differences in copy number (copy number variation; CNV) of genes and repeats, presence-absence variation (PAV), various forms of chromosomal change such as inversions (Huang and Rieseberg, 2020) and translocations, and homoeologous exchanges (HEs; see Mason and Wendel, this issue). In fact, the two or more co-resident genomes of allopolyploids almost certainly contain structural variations, which have been shown to accumulate in all plant genomes studied (Saxena et al., 2014; Fuentes et al., 2019; Gabur et al., 2019; Schiessl et al., 2019). Thus, divergence of two diploids is accompanied by the natural accumulation of structural differences, which then become combined in a common nucleus during allopolyploidization.

The scale of structural variation within and among plant species represents an extraordinary discovery of the genomics era; that is, rather than SVs being rare, one-off mutants, plant genomes appear to be rife with this form of diversity, so much so that "reference genomes" are now widely thought of as providing only a snapshot of the "pangenome" that actually characterizes a species (or group of species) (Golicz et al., 2016; Danilevicz et al., 2020). Different rice lines, for example, collectively contain at least 1.5 million SVs (Fuentes et al., 2019), and even in a relatively limited sampling of 19 maize inbreds and 14 teosintes, approximately 4000 *genes* experience either CNV or PAV (Swanson-Wagner et al., 2010). Similarly, about 9% of the 26,000 genes in a sampling of 80 *Arabidopsis* lines are missing in at least one line (Tan et al., 2012), and in a sampling of 115 cucumber lines, variation caused by SVs affects 1676 genes (Zhang et al., 2015). Accordingly, there is every reason to suspect that CNVs and PAVs can affect phenotypes and be relevant to natural selection and adaptation.

Evidence in support of this assertion now abounds, some from natural systems (Winzer et al., 2012; Flagel et al., 2014), but mostly from the crop literature (e.g., Wang et al., 2015; Zhang et al., 2015; Gabur et al., 2019; Schiessl et al., 2019). Importantly, from the standpoint of the present article, examples from polyploid crops are accumulating (Gaeta et al., 2007; Gabur et al., 2019; Schiessl et al., 2019), including oilseed rape (*Brassica napus*), potatoes (*Solanum tuberosum*), and bread wheat (*Triticum aestivum*). Traits for which SVs have been implicated as causative include some that are readily envisioned to be responsive to natural selection, such as flowering time and frost tolerance (Gabur et al., 2019; Schiessl et al., 2019). Given the scale and scope of SVs in all plants studied to date, it seems likely that their role in adaptive processes will increasingly be recognized as important, and especially in the adaptation and diversification of nascent allopolyploids, which are forged from the merger of two genomes that bring to the union differing suites of SVs.

In addition to small scale SVs affecting copy number or presence/absence of genes, exons, small repeats and the like, larger structural mutations abound in polyploids, resulting from

processes such as reciprocal and non-reciprocal homoeologous exchanges (HEs) (Mason and Wendel, this issue). These mutations, which affect genomic regions ranging from smaller telomeric regions to interstitial segments to entire chromosome arms, have the capacity to simultaneously alter genic and non-genic PAVs and copy number dosages on a massive scale. Classically recognized as homoeologous translocations or transpositions, the genomics era and the successful sequencing of polyploid plants ushered in an increasing realization that HEs represent a fundamental mechanism of allopolyploid genome evolution and for generating diversity. Recent illustrative examples include peanuts (*Arachis hypogaea*; Bertioli et al., 2019; Zhuang et al., 2019), *Tragopogon* (Chester et al., 2012), quinoa (*Chenopodium quinoa*; Jarvis et al., 2017), *Brassica* (Hurgobin et al., 2018; Samans et al., 2018), tobacco (*Nicotiana tabacum*; Chen et al., 2018) and allopolyploid rice (constructed from *Oryza sativa subsp. indica* × *subsp. japonica*; Sun et al., 2017; Li C. et al., 2019). Increasing evidence suggests that in many systems HEs may occur genome-wide and be sequential, ongoing, and variable in size. In principle, this process can generate a limitless pool of genetically variable progeny over time, with each genomic combination carrying its own particular suite of chromosome segment copy numbers. One might imagine that the immense range of genomic diversity generated by an ongoing process of HE would be paralleled by phenotypic diversity, and thus it might represent a potent force for adaptation and evolutionary change in polyploids.

Hints that HEs might generate selectively relevant phenotypic diversity during allopolyploid evolution extend back at least to 2007 (Gaeta et al., 2007), and their association with PAVs has now been clearly established (Sun et al., 2017; Hurgobin et al., 2018; Li C. et al., 2019). In *Brassica*, HEs affect flowering time, disease resistance, and glucosinolate metabolism (Hurgobin et al., 2018). In rice allopolyploids, genome-wide gene expression and methylation states are massively altered by HEs, which also are associated with diverse phenotypes (Sun et al., 2017; Li C. et al., 2019; and unpubl.); most impressively, these outcomes arise in even the first few generations of selfing from a single founder following artificial allopolyploid synthesis (e.g., Chester et al., 2012; Sun et al., 2017; Hurgobin et al., 2018), demonstrating that HEs likely are a potent force for evolutionary novelty following allopolyploid speciation, at least in some groups.

Whole-Genome Effects

A poorly understood but undoubtedly significant dimension of polyploidy concerns the evolutionary relevance of the myriad cascading effects set in motion by the doubling (in the case of autopolyploidy) or summing of genome sizes into a common nucleus. This additivity, by itself, is known to trigger diverse regulatory alterations in gene expression, translation, biosynthesis of metabolites and structures, cells sizes and shapes, organ size, physiology, and almost any aspect of plant development that one studies. Even at the level of the genome, recent studies using chromosome conformation capture, or Hi-C (Grob, 2020), have demonstrated that genome merger and WGD also dramatically alter the positional relationships and associations among chromosomes in the nucleus. A case in point is the remarkable study by Wang et al. (2018), who showed that in allopolyploid cotton, the suites of Topologically Associated Domains (TADs) within and between chromosomes are significantly altered by allopolyploidy relative to the diploid progenitors, and that some homoeologous chromosomal regions become spatially associated whereas others do not. Similarly, Zhang et al. (2019) recently showed that in autopolyploid *Arabidopsis thaliana*, chromosome doubling led to an increase in interchromosomal interactions and decreased association of more closely adjacent intrachromosomal sites. The effects of these types of spatial and organizational alterations on gene expression dynamics and all of the downstream reverberations that lead to phenotypes are largely unknown. Yet some data are beginning to close this circle; Zhang et al. (2019), for example, also showed that the altered chromatin interactions were associated with specific changes in gene expression and histone modifications that might affect phenotypes, including for the key flowering regulator *Flowering Locus C*.

Above the level of the genome, a burgeoning but highly fragmented literature exists bearing on one or more aspects of the suite of scaling and stoichiometric changes set in motion by polyploidy, as elegantly and comprehensively reviewed recently by Doyle and Coate (2019). Notwithstanding the examples cited in their review and in other parts of the present perspective piece, the fact remains that nearly all polyploidy-induced phenotypes, structural, metabolic, or physiological, that one might consider phenotypically relevant to adaptation, in either natural settings or in domesticated plants, represent emergent, downstream features of complex cascading networks of genic, regulatory and biosynthetic programs. As such, it is perhaps unsurprising that understanding the "genotype to phenotype" (G-to-P) mapping equation remains elusive for nearly all traits distinguishing diploid from allopolyploid congeners. Partial solutions to the G-to-P equation are provided for some complex traits using tools from multiple "omics" and scales. Examples including ploidy-related invasiveness potential in goldenrods (Wuet al., 2020), growth rate and phenotypic traits such as cell size and cell wall composition among *A. thaliana* plants having different levels of autopolyploidy (Corneillie et al., 2019), and even the rapid rise to global prominence of angiosperms following ancient WGD events (Simonin and Roddy, 2018). As noted by Doyle and Coate in citing Don Levin nearly 40 years ago (1983), "*the role of polyploidy per se in the development of evolutionary novelty remains one of the outstanding questions in flowering plant evolution*". At the same time, Doyle and Coate offer promising ideas for a research agenda directed at this question, and as pointed out elsewhere in the present review, the tools and technologies available today are getting us closer to this holy grail. Some of these are alluded to in the following section.

RESPONSES AT THE GENE EXPRESSION AND REGULATORY LEVEL

Duplicate Gene Expression Evolution

One of the revelations of the genomics era, unsurprising in hindsight, is that the merger of two or more differentiated

genomes, in the case of allopolyploidy, or the doubling of a single genome, in the case of autopolyploidy, causes massive, genome-wide alteration in gene expression patterns. These alterations accompany both hybridization and polyploidy separately, encompassing both immediate or short-term consequences of genome merger as well as evolved responses that arise over thousands to millions of years subsequent to WGD (Flagel et al., 2008). This temporal partitioning is useful when thinking about the evolutionary relevance of gene expression evolution, as it seems likely that many or perhaps most evolved responses were in fact enabled by relatively ancient polyploidy events, but that the adaptive signatures of such evolutionary change remain either obscure or are no longer evident. One example of this might be the well-known "radiation lag-time" hypothesis (Schranz et al., 2012; but see Tank et al., 2015; Landis et al., 2018), which was proposed to explain an observation of a delay, or lag phase, between inferred ancient polyploidy events and diversification. Irrespective of the effects of polyploidy on net diversification rates, the notion that doubled genomes may "have time" to generate adaptive phenotypes represents an important idea for understanding the evolutionary potential of polyploidy. Phrased alternatively, unlike the case in conventional diploids, the additive complement of genes, regulatory elements, and other genomic components that comprise a nascent polyploid (auto- or allo-) represent a vast storehouse of raw material for later, and perhaps much later adaptive responses to environmental change, niche expansion, or any other biotic or abiotic evolutionary opportunity.

Numerous phenomena and analytical frameworks are encompassed by the terms "expression evolution" or "regulatory evolution," ranging from those focused on single pairs of duplicated genes (homoeologs) to others involving entire networks of coexpression for hundreds to thousands of duplicated genes. Many of these topics have been amply reviewed (Yoo et al., 2014; Soltis et al., 2015, 2016; Wendel, 2015; Panchy et al., 2016; Van de Peer et al., 2017; Bird et al., 2018; Cheng et al., 2018; Wendel et al., 2018). Here our attention is directed at the connections, known or suspected, between expression or regulatory evolution and plant phenotypes.

An early and illustrative example of differential homoeolog expression ("homoeolog bias"; Grover et al., 2012), is from allopolyploid (AD genome) cotton (*Gossypium*), where it was shown that the two co-resident, alternative homoeologs (A, D) had differential contributions to the total transcript pool, and that this homoeolog ratio varied widely among genes (Adams et al., 2003). In the most extreme cases, reciprocal silencing of alternative homoeologs was observed for different tissues; for example, in petals and stamens, only the D homoeolog was expressed, whereas in ovary walls of the same flower, only the A homoeolog was expressed. Whereas not all genes in every allopolyploid exhibits this degree of homoeolog bias, the phenomenon itself is an ubiquitous feature of polyploids (Grover et al., 2012; Yoo et al., 2014), and it may be that there are few if any homoeologous gene pairs in any species that contribute equally to the transcript pool in all tissues. Thus, homoeolog bias appears to be a rule rather than an exception. Yet differing transcript ratios may not be evolutionarily meaningful in terms of protein function if the genes encoded by the two homoeologs are functionally equivalent and do not differ in expression domains. Accordingly, evolutionary relevance is more likely in situations where different "function" has been inferred for two homoeologs, either prior to hybridization and WGD or as an evolved feature subsequent to polyploidization. These two cases are quite different in terms of our understanding of the timing of evolutionary divergence, that is, whether homoeolog functional differences arose at the diploid or the polyploid level. Distinguishing between these two cases often is not possible because the progenitor diploids are unknown or extinct, and in this respect a number of plant genera have become particularly useful models (Soltis et al., 2016), including *Gossypium, Nicotiana, Arabidopsis, Tragopogon, Senecio, Glycine, Brassica, Spartina, Aegilops-Triticum*.

Functional divergence and other forms of neofunctionalization of homoeologs involves the acquisition of novel expression domains, interactions, or protein function following their merger in a common polyploid nucleus, and it is the duplicate gene outcome of most interest to adaptation. Other possible fates of duplicated genes, not wholly separable from neofunctionalization, include subfunctionalization, where ancestral aggregate expression space or function is partitioned developmentally or in a tissue-specific fashion following polyploidy, non-functionalization (mutational loss of one or the other homoeolog), dosage subfunctionalization (Gout and Lynch, 2015), and several other related outcomes. These and other possibilities regarding the evolutionary fates of gene duplication have been extensively reviewed (e.g., Conant and Wolfe, 2008; Flagel and Wendel, 2009; Panchy et al., 2016; Cheng et al., 2018). While these categorizations are conceptually useful, most empirical examples defy a simple characterization, as divergence often involves multiple steps and various combinations of duplication, loss, subfunctionalization and neofunctionalization.

An excellent example of neo- and subfunctionalization concerns a pair of genes in the Brassicaceae duplicated by a polyploidy event about 23 MYA (Liu and Adams, 2010). The two paralogs SHORT SUSPENSOR (SSP), involved in paternal control of zygote elongation in *A. thaliana*, and Brassinosteroid Kinase 1 (BSK1), involved in brassinosteroid signal transduction, have diverged in function, with BSK1 retaining its ancestral role in hormone signaling. SSP, however, diverged following duplication to acquire a role in zygote elongation, while losing its plesiomorphic role in signal transduction through loss of its kinase domain. A second illustrative example concerns the fate of duplicated MADS-box transcription factors in columbine (*Aquilegia*) flowers (Sharma and Kramer, 2013). Duplication of ancestral APETALA3 genes in the Ranunculaceae appears to have been followed by a combination of neo- and subfunctionalization of different paralogs, while contributing to the evolution of a novel floral whorl where the innermost stamens have been converted to sterile staminodes. A third example concerns the free-threshing Q locus in hexaploid wheat (Zhang et al., 2011), which encodes an AP2-type transcription factor. This locus has a complex history of duplication and differential paralog loss among diploid wheat lineages during the evolution

of *Triticum/Aegilops*. Diverged paralogs were reunited in a common nucleus with polyploid wheat formation, and this was followed both by protein evolution in one homoeolog and pseudogenization/subfunctionalization of other homoeologs. This example entails a complex combination of ancient paralogy, non-functionalization, reunion of divergent paralogs, and interaction of subfunctionalized homoeologs. These specific examples of post-duplication evolutionary divergence differ greatly in their timing, underlying molecular bases, and ecological settings. It remains to be seen whether commonalities in any of these attributes will emerge in the future as additional examples are revealed.

In addition to examples involving duplicated genes connected to specific phenotypes, broader surveys of patterns of gene retention following polyploidy have implicated certain classes of genes in morphological innovation or in adaptation to various environmental conditions. Transcription factors, for example, have been shown to be preferentially retained following ancient polyploidy events (De Smet et al., 2013; Li et al., 2016), and many of these retained ancient homoeologs later become modern paralogs with differing roles in regulatory development and plant morphology (reviewed in Rensing, 2014). Schilling et al. (2020) studied 201 wheat MIKC-type MADS-box genes, showing both preferential retention and even gene family expansion for genes involved in adaptation to different environmental conditions, abiotic and biotic stresses, and flowering time. These and other studies indicate that transcription factor retention and divergence appears to be an important aspect of polyploid diversification. Insight into the nature of this preferential retention was recently provided by Panchy et al. (2019), who showed that rapid evolution of *cis* binding sites generates novel TF expression patterns that lead to subfunctionalization and neofunctionalization as well as complex combinations of new and ancestral expression states.

Additional insights into the evolution of duplicated genes has emerged from studies of suites of genes tracing to a common progenitor genome, rather than a common class or category of gene. An exemplar study in this respect is that of De Smet et al. (2017) who demonstrated a pattern of coordinated gene expression following ancient polyploidy in *A. thaliana*. In this case a suite of 92 homoeologous gene pairs were identified which shared a common pattern of tissue-specific gene expression, but with the two homoeologous suites being partitioned such that one was expressed mostly in aerial tissues while the other was expressed predominantly in roots. This remarkable example of coordinated homoeolog evolution among diverse sets of genes has parallels to recent discoveries from using coexpression network approaches, discussed below.

Altered Epigenetic Landscapes

The observation that global patterns of gene expression are dramatically altered by a change in ploidy (above) led to the supposition that at least part of this response must have an epigenetic basis. This supposition was confirmed by studies of synthetic polyploids and interspecific F1 hybrids, plants in which there has been virtually no chance for mutational effects yet which exhibit massive gene expression modification (Adams et al., 2003, 2004; Flagel et al., 2008; Yoo et al., 2014; Song and Chen, 2015; Ding and Chen, 2018). This regulatory rewiring of the transcriptome likely has numerous combinatorial causes stemming from altered cell and nuclear volumes (Doyle and Coate, 2019), biochemical and biophysical stoichiometric disruptions (Bottani et al., 2018; Hu and Wendel, 2019), and changed *cis* and *trans* controls on gene expression (Bao et al., 2019). All of these phenomena are related to various forms of epigenetic modification and chromatin remodeling, including changes in DNA methylation and various histone modifications, for both autopolyploids and allopolyploids (Song and Chen, 2015; Ding and Chen, 2018).

A beautiful example of the relevance of these epigenetic phenomena to novel plant phenotypes associated with allopolyploidy concerns flowering time in domesticated forms of allotetraploid (AD genome) cotton (Song et al., 2017). Wild forms of both *Gossypium barbadense* and *G. hirsutum* contain hypermethylated forms of both homoeologs (A, D) of CONSTANS-LIKE 2 (COL2) and are photoperiod sensitive. As a consequence of domestication, however, DNA methylation was lost for the D homoeolog, leading to higher expression of COL2D and the all-important photoperiod insensitivity that allowed cotton production to thrive outside of the subtropics. A second example concerns circadian clock genes in *Arabidopsis* hybrids and allotetraploids (Ni et al., 2009), where histone modifications were linked to increased biomass, vigor, and starch accumulation. A third example of an epigenetically mediated plant trait accompanying polyploidy, recent or ancient, is the possible epigenetic neofunctionalization of a parentally imprinted polycomb group protein in grasses (Dickinson et al., 2012), which may have contributed to the evolution of the globally important large endosperm found in cereals. Finally, Lu et al. (2020) recently observed that the polyploid populations of *Solidago canadensis*, which have a more southerly distribution than their diploid counterparts, and which have a lower freezing tolerance, had lower expression levels but more copies of a key gene (ScICE1) involved in freezing tolerance. In this case the authors suggested that promoter DNA methylation has repressed expression, leading to polyploid adaptation accompanying range expansion.

It seems likely that we have only just begun to understand the relationships among epigenetic responses to polyploidy and novel phenotypes or adaptation, but given the scale and the scope of both polyploidy and its unavoidable epigenetic consequences, it seems probable that many new examples will soon emerge.

Finally, we note the potentially important observation that reciprocal homoeolog silencing, as described for different organs of the same plant in *Gossypium* (Adams et al., 2003, 2004), can arise immediately upon polyploid formation. Moreover, the same tissue-specific expression pattern has been observed in natural allopolyploids 1–2 million years following initial polyploid formation. This suggests the tantalizing possibility that there has been stable maintenance of an epigenetic mutation since the initial formation of allopolyploid *Gossypium*. Adams et al. (2004) noted the important temporal dimension of this epigenetic suppression of gene expression, suggesting that epigenetic subfunctionalization may provide a selective constraint favoring

duplicate gene retention, in that both homoeologs would be necessary to enable expression in alternative tissues. To the extent that this is true, epigenetic subfunctionalization might provide a latent reservoir of hundreds to thousands of genes, which "become exposed to an evolutionary filter only after additional epigenetic and genetic evolution" (Adams et al., 2004, p. 2225) perhaps thousands to millions of years later. The scale of the phenomenon of homoeolog bias and reciprocal silencing, as noted above, suggests that epigenetic maintenance via subfunctionalization may prove to be a significant facet of polyploid evolution.

Small RNA Duplication and Divergence

Another genomic facet of hybridization and polyploidy is the attendant combining of diverged populations of small RNAs (sRNAs), including microRNAs (miRNAs) and several classes of small interfering RNAs (siRNAs) (Ng et al., 2012; Wendel et al., 2016; D'Ario et al., 2017). Each of these classes of 21 to 24 nucleotide RNAs has a suppressive effect on either gene or transposable element (TE) expression, and all are subject to the novel *trans*-regulatory controls established by genome merger. As such, the combination of two populations of diverged sRNAs has the potential to change patterns of gene expression, TE activity, and all of the developmental programs and phenotypes that might result from this altered regulatory environment. Micro RNAs, for example, are known to be important players in stress responses and other forms of ecological adaptation (Song et al., 2019; Yu et al., 2019), which raises the possibility that these small regulatory molecules might facilitate neofunctionalization of duplicated stress-related and other signaling genes (Palacios et al., 2019).

A growing literature attests to the effects of polyploidy on expression of small RNAs (reviewed in Ng et al., 2012; Wendel et al., 2016). As with protein-coding genes, in most studies at least some sRNAs are non-additively expressed (e.g., Li et al., 2014; Xie and Zhang, 2015; Cavé-Radet et al., 2019; Palacios et al., 2019; Wei et al., 2019), being subject to novel *trans* regulation. sRNA non-additivity in polyploids may be unsurprising in that this also has been observed within species as well, as in maize hybrids (Crisp et al., 2020). For miRNAs, this non-additivity will affect downstream gene regulation, so in principle non-additivity should have consequences for plant phenotypes; for siRNAs, non-additive siRNA expression could, in *trans*, affect epigenetic silencing or activation of TEs (transposon elements), which could also exert effects on neighboring protein-coding genes and eventually cause phenotypic alterations (see below). There are, however, surprisingly few examples of where sRNA divergence has been causally connected to novel traits in polyploids. In cotton allopolyploids, for example, homoeologous MYB2 transcription factors, regulated by miR828 and miR858, are important regulators of cotton fiber development (Guan et al., 2014). These miRNAs have functionally diverged with respect to their targeting preferences, and are inferred to contribute in a novel manner to polyploid cotton fiber growth. It seems probable that additional examples will soon emerge from functional analysis of the omnipresent alterations in sRNA expression in polyploids.

Altered *Cis* and *Trans* Relationships

Allopolyploidy entails the merger of two suites of partially diverged *cis*- and *trans*-regulatory elements, and as such gene expression is expected to be altered due to the several new forms of regulatory interactions in the polyploid nucleus (Bottani et al., 2018; Hu and Wendel, 2019). These expectations recently were illuminated in an experiment where reciprocal F1 hybrids were constructed between cultivated and wild accessions of the allotetraploid cotton species *G. hirsutum* (Bao et al., 2019). Although the goal was to understand the nature of the domestication process, the study revealed a surprisingly high level of *trans*-regulatory control of gene expression (54–64%), higher than observed in comparable studies in diploids. Bao et al. (2019) proposed the explanation that with the onset of allopolyploidy, *trans* factors throughout the genome instantaneously acquire duplicated homoeologous suites of *cis* elements with which to interact. This aspect of allopolyploidy generates extensive and novel *cis-trans* interactions, especially for *trans* variants. As noted by Bao et al. (2019), "This phenomenon of enhanced *trans* regulatory evolution may be a general and previously unrecognized feature of polyploidy, perhaps helping to explain evolutionary novelty in recently formed allopolyploid plants." It seems probable that in the next few years specific examples will emerge where novel phenotypes are causally connected to these new regulatory interactions.

COMBINATORIAL GENOMIC AND EXPRESSION-LEVEL RESPONSES

Novel Cytonuclear Combinations

Allopolyploid formation not only results in the merger of two nuclear genomes, but because it is directional there is a pollen donor and an ovule donor. The latter is the source of the cytoplasm with its mitochondria and plastids in perhaps 80% of angiosperms (Corriveau and Coleman, 1988), and accordingly the nascent allopolyploid and all descendant lineages have organellar genomes tracing to the maternal diploid parent. This directional asymmetry also is accompanied by a genic imbalance, as polyploidization results in a doubling of nuclear gene content (for a tetraploid) with a more uncertain quantitative effect on cytoplasmic genome number (Sharbrough et al., 2017; Doyle and Coate, 2019). Because organellar genomes diverge during diploid divergence, in many of the same ways as those discussed above for nuclear genomes, and because many aspects of plant physiology and development involve finely tuned integration of plastidial and mitochondrial processes with anterograde and retrograde nuclear-organellar signaling, there has been a long interest in the evolutionary dimension of this cytonuclear relationship (reviewed in, e.g., Bock et al., 2014; Sharbrough et al., 2017; Fishman and Sweigart, 2018). This work includes observational, experimental, and statistical evidence bearing on the fitness of different cytonuclear combinations within and between species (Bock et al., 2014; Case et al., 2016; Roux F. et al., 2016) as well as a vast literature on cytonuclear incompatibility (Fishman and Sweigart, 2018) and the important topic of

cytoplasmic male sterility in crop plants (Chen et al., 2017). That specific cytonuclear combinations can affect plant phenotype has now been firmly established. This was recently convincingly demonstrated by Flood et al. (2020) for 1,859 phenotypes in all possible cybrid combinations among seven *A. thaliana* lines.

With respect to hybridization and polyploidy, mounting evidence suggests that cytonuclear molecular coevolutionary responses will be common. One source of evidence is "global"; for example, nuclear genes encoding organellar processes are preferentially restored to single-copy status following polyploidy events (De Smet et al., 2013), presumably to restore "normal" stoichiometric relationships with organellar-encoded proteins. Another example is that of the D-genome wheat species *Aegilops tauschii*, which has a homoploid hybrid origin and which has preferentially retained genes from its maternal A-genome ancestry for nuclear genes that encode cytonuclear enzyme complexes (Li N. et al., 2019). At a more granular level, Gong et al. (2012, 2014) studied RuBisCO in allopolyploids of five different model genera, *Arabidopsis, Arachis, Brassica, Gossypium,* and *Nicotiana*, demonstrating that paternal copies of the nuclear-encoded small subunit (*rbcS*) experience gene conversion such that their sequences became maternal-like, and also that there was preferential expression of the maternal *rbcS* copies. These putative coevolutionary responses were not found, however, in some other genera, e.g., *Tragopogon* (Sehrish et al., 2015) and *Cucumis* (Zhai et al., 2019).

A potential connection to phenotype, and an example that integrates several of the mechanisms discussed in the present review, is from Wu et al. (unpublished), who studied synthetic allopolyploids synthesized from reciprocal crosses between rice *O. sativa* subsp. *sativa* and subsp. *japonica.* Each generation of selfing was accompanied by multiple homoeologous exchanges (see earlier discussion), collectively affecting all members of the chromosome complement, such that after four generations of selfing the resulting individuals were genomic mosaics of the two founding parents. Importantly, some genomic regions were preferentially and reciprocally biased with respect to maternal vs. paternal progenitor in that they were repeatedly associated with only one of the two parental cytoplasms, whereas other chromosomal regions were exclusively maintained as heterozygotes, suggesting hetero-cytonuclear interactions.

As with many of the other phenomena discussed here, the foregoing synopsis suggests an adaptive dimension to cytonuclear interactions that may be set in motion through hybridization and genome doubling.

Effects on Transposable Element Activity

Transposon elements are mobile genetic elements that account for a large but variable fraction of virtually all eukaryotic genomes, including plants. As an example, LTR-retrotransposons, which together with miniature inverted-repeat transposable elements (MITEs) constitute the most prevalent and active class of TEs in plants (Casacuberta and Santiago, 2003), account for only 2.5% of the small and compact genome of *Utricularia gibba* (Ibarra-Laclette et al., 2013) but 90% of the gigantic genome of *Fritillaria* spp. (Ambrožová et al., 2011). In fact, TE proliferation and TE elimination by transposon-mediated recombination and deletion-biased double strand break (DSB) repair are two of the primary drivers of genome expansion and shrinkage during land plant evolution (Pellicer et al., 2018). These processes can play out saltationally to generate large changes in genome size even over short evolutionary time scales. As an example, the proliferation of a few retrotransposon families explains the doubling of the genome size of *Oryza australiensis* relative to rice (Piegu et al., 2006). Similarly, the 3-fold difference in genome sizes among different diploid cotton (*Gossypium*) species reflects the differential dynamics of TE proliferation and clearance since these species shared a common ancestor (Hawkins et al., 2009). Even within species there can be extensive TE polymorphism (Carpentier et al., 2019; Noshay et al., 2019). Thus, the TE component of plant genomes is highly variable in plants, even among closely related species and often within species.

In the context of the present review, the merger of two different genomes will inevitably combine two different TE populations, which is of particular relevance because the presence and mobility of TEs impact genomes in many ways. First, their transposition induces new insertions, which in most cases will be selectively neutral or slightly deleterious, but in other cases could provide a selective advantage (Arkhipova, 2018). Second, their repetitive nature offers numerous pairs of sequences that can recombine, and accordingly, TEs are a major source of structural variants, including genic CNV and PAV, as discussed earlier (see ***Structural diversity***, above, see also Fuentes et al., 2019). And third, TEs are a rich source of new genes and gene functions and can directly or indirectly regulate gene expression (Lisch, 2013). Indeed, TEs are an important source of promoters and transcriptional regulatory elements. Transcription is the first step of transposition, and TEs contain internal promoters to facilitate their own expression. The insertion of TEs within or close to genes can therefore alter the expression of neighboring genes by providing additional transcription factor binding sites or alternative promoters and splicing signals, a phenomenon frequently found in both animal (Chuong et al., 2017) and plant genomes (Lisch, 2013).

Transposon elements accumulate in certain regions of the genome (e.g., centromeres and pericentromeric regions) where they can fulfill important structural functions, as for example supporting the specification and function of centromeres (Lermontova et al., 2015). But this non-homogeneous accumulation in chromosomes also impacts genic evolution. In filamentous fungi, the "two-speed genomes" concept has been developed to describe the concentration of the fast-evolving virulence effectors in TE-rich compartments of the genome (Dong et al., 2015). The formation of TE islands and differentiated genomic regions showing distinct evolutionary rates could allow for the rapid evolution required for adaptation to new environments while preserving the basic genic machinery (Schrader and Schmitz, 2019), as has been shown for the invasive ant *Cardiocondyla obscurior* (Schrader et al., 2014). Although this phenomenon has not been described as such in plants (Lanciano and Mirouze, 2018), there are accumulating examples of TE impact on the evolution of different plant resistances to

biotic and abiotic stress, such as disease resistance in pepper (Kim et al., 2017) and aluminum resistance in a wide range of land plants (Pereira and Ryan, 2019), and it has long been known that plant disease resistant genes frequently concentrate in resistance gene clusters which are also rich in TEs (Richter and Ronald, 2000). Moreover, the differential distribution of genes with respect to their age or function in genomic compartments defined by a different TE content has recently been shown in plants, including tomato (Jouffroy et al., 2016) and melon (Morata et al., 2018b), suggesting that TE-rich compartments may facilitate rapid adaptation in plants. In addition, TEs can also modify the local recombination rate along chromosomes. Indeed, a recent study in natural populations of *Arabis alpina* has shown that TEs can create linkage disequilibrium blocks defining adaptive loci that concentrate environment-responsive genes (Choudhury et al., 2019).

Given their ubiquity, extraordinary variability even within species, and propensity to function as genomic architects, it is likely that TEs are important internal drivers of plant evolution and adaptation. However, TEs also pose a hazard for genome integrity, as they are a cause of potentially deleterious mutations and chromosome instability. For this reason, TE activity is highly repressed by different mechanisms, the most important being epigenetic silencing driven by DNA methylation (Kim and Zilberman, 2014; Zhang H. et al., 2018). TEs are the main target of epigenetic silencing marks in the genome, and TE distribution closely matches that of DNA methylation (El Baidouri et al., 2015). The equilibrium between efficient silencing mechanisms to control TEs, and the escape of some TEs from this control under particular circumstances, allows for TE maintenance and genome plasticity while maintaining genome integrity. TEs are mostly quiescent and are only activated in particular developmental stages (Martínez and Slotkin, 2012) or under stress (Galindo-González et al., 2017). This activation of TEs by stress, and probably also during development, likely is the result of the combination of the presence of specific activator sequences in TE promoters (Galindo-González et al., 2017), and the alleviation of epigenetic silencing in these situations (Gutzat and Mittelsten Scheid, 2012). The activation of TEs under stress allows for the generation of new variability in situations to which the genome is not necessarily well-adapted. In addition, the insertion of TEs with stress-related promoters close to genes could result in the stress-related expression of a new set of genes offering new possibilities for adaptation to new environmental conditions. Interestingly, TE integrations are frequently not random, and it recently has been shown that some LTR-retrotransposons preferentially target environmentally responsive genes (Quadrana et al., 2019), generating new genetic or epigenetic variability that could facilitate rapid adaptation to new environments.

Although LTR-retrotransposons are the most obvious candidates for altering adjacent gene expression, other TEs also have this potential. In particular, MITEs, which are present in high copy number and are enriched in genic regions (Casacuberta and Santiago, 2003), frequently contain transcription factor binding sites (Morata et al., 2018a). In addition, TEs of different types have been shown to alter gene splicing upon insertion. As an example, an Helitron insertion into a host susceptibility factor gene causes its alternative splicing leading to resistance to maize rough dwarf disease (Liu et al., 2020). Moreover, as noted above, TEs are the main target of epigenetic silencing, and therefore, the insertion of a TE within or close to a gene may bring the silencing machinery to this gene altering its expression (Zhang H. et al., 2018). In addition to bringing new promoters and promoter elements, TEs can also be the source of small RNAs that regulate gene expression (McCue and Slotkin, 2012), and in particular that of defense-related genes (Poretti et al., 2020), and it has been recently shown that they can also contribute long non-coding RNAs that regulate plant stress responses (Lv et al., 2019). TEs are therefore an important force that facilitates plant, and also animal (Rech et al., 2019) and yeast (Esnault et al., 2019), adaptation to stress. This confirms McClintock's revolutionary hypothesis on the role of mobile genetic elements in overcoming the threat of environmental shock by reorganizing the genome (McClintock, 1984).

McClintock's proposal regarding the "shock" that follows genome merger has been amply evidenced by data showing that plant hybridization and polyploidization frequently trigger TE activation (Vicient and Casacuberta, 2017). Transcriptional activation of TEs has been reported in synthetic *Arabidopsis* polyploids (Madlung et al., 2005), wheat amphiploids (Kashkush et al., 2003), allopolyploid coffee (Lopes et al., 2013), and in rice lines derived from introgressive hybridization with *Zizania latifolia* (Wang et al., 2010), among others. This activation of TEs in hybrids and polyploids has been shown to be accompanied in many cases by a modification of the siRNAs that target TEs (Springer et al., 2016). Moreover, TE mobilization and increase in copy number has also been reported for different hybrids and polyploids, including tobacco (Petit et al., 2010), wheat (Yaakov and Kashkush, 2012) and *Brassica* (Sarilar et al., 2013) allopolyploids, in *Biscutella laevigata* autopolyploids (Bardil et al., 2015), and in sunflower hybrids (Kawakami et al., 2010). However, in other allopolyploids or interspecific crosses no evidence of an increase of TE content was shown. For example, no changes were observed in TE regulation after an interspecific cross between *A. thaliana* and *A. lyrata* (Göbel et al., 2018), and no TE burst was detected after polyploidization of *A. arenosa* (Baduel et al., 2019). Moreover, an increase of TE-related siRNAs was recently reported in *Spartina* allopolyploids, suggesting a strengthening of TE repression accompanying polyploidization (Cavé-Radet et al., 2019).

An important dimension of TE mobilization following genome merger is that this varies among TE families, and the same TE may proliferate in some polyploids while being eliminated in others. As an example, *Sabine* retrotransposons proliferated following polyploidy in some *Aegilops* (wheat) polyploids while being eliminated in others (Senerchia et al., 2014), even while other retrotransposon families (*BARE1*, *Romani*) experienced a more uniform proliferation. Different TE families are regulated dissimilarly and their responses to the methylation and siRNA changes that accompany genome merger may vary accordingly. At present there is little understanding of this extraordinary variation in TE responses to polyploidy, but we can imagine that this reflects an equivalent level of diversity

with respect to divergence in progenitor TE populations and their repression and activation dynamics among parental diploids in each genus. In this respect, it has recently been shown that there is a correlation between TE mobilization in *Nicotiana* allopolyploids and the quantitative imbalance in parental TE loads (Mhiri et al., 2019). But even in the absence of TE activation, polyploidy may result in an increase in TE content due to relaxed purifying selection at duplicated loci (Ågren et al., 2016; Baduel et al., 2019).

In summary, merging two different genomes, with or without polyploidization, will combine in a single genome two different TE populations, together with the siRNAs that target them, and this may result in changes in the epigenetic modifications at TEs and neighboring genes and in the regulation of genes and TEs.

Given the prevalence of TEs, their frequent association with genes, and their potential for insertional mutagenesis following genome merger, it seems probable that they are important players in the creation of novel traits and adaptation in polyploids. As explained above, the importance of TEs in creating phenotypic variability in plants is well established (Lisch, 2013), and examples of the role of TEs in creating new adaptive alleles are slowly accumulating. For example, TEs have been recently shown to create adaptive alleles that modify flowering time (Huang et al., 2017; Liang et al., 2019; Niu et al., 2019), facilitate local climate adaptation (He et al., 2018), and trigger new responses to biotic (Poretti et al., 2020) and abiotic (Lv et al., 2019) stresses, sometimes by facilitating the formation of complex biosynthetic pathways (Xu et al., 2017). However, direct evidence of TEs generating new adaptive phenotypes as a consequence of the merging of two genomes remains scarce. As one example, for TE families targeting environmentally responsive genes, activation may introduce target variability at these loci, potentially facilitating rapid adaptation (Quadrana et al., 2019). An added dimension to this scenario is that relaxed purifying selection in polyploids (due to duplication) can also result in accumulation of TEs close to environmentally responsive genes (Baduel et al., 2019). The combination of TE activation and the relaxed purifying selection that frequently accompanies genome merger provides a powerful mechanism for the generation of novel allelic combinations for stress-related or other adaptive responses. We note that in principle adaptation may be facilitated by selection on even a single TE insertion affecting a single gene, as illustrated by the beautiful example of the peppered moth industrial melanism mutation (Hof et al., 2016), or it might involve the evolution of complex pathways, as has been shown for the nicotine synthesis in *Nicotiana* (Xu et al., 2017).

Disrupted/Altered Regulatory Networks

It may be useful to consider that each of the many mechanisms introduced in the foregoing sections, from simple genic or regulatory SNPs through novel transposable element activity, ultimately shape phenotypes via their propagation through complex networks of gene regulation, transcription and translation, and higher order biochemical, physiological and biosynthetic processes (Gottlieb, 1984). To this extent, nearly all selectively relevant phenotypes likely represent emergent properties resulting from high-level multidimensional, interconnected meshworks of lower level "omic" processes. We expect that we are entering a period during which this omics-enabled view of adaptation and evolutionary change will rapidly expand, concomitant with the application of a suite of enabling technologies to model experimental and natural systems.

A foreshadowing of this form of evolutionary exploration is offered by the development of coexpression network analysis, in which entire transcriptomes are interrogated for patterns of genic and "modular" coexpression, or lack thereof. The logic for this rests on the assumption stated above, i.e., that genes work in concert rather than in isolation to generate phenotypes. Many aspects of genic coexpression network analysis have been amply reviewed (e.g., Serin et al., 2016; Emamjomeh et al., 2017; Ruprecht et al., 2017b; Contreras-Lopez et al., 2018; Joehanes, 2018; Rao and Dixon, 2019); here our attention is focused on applications involving adaptation following genome duplication and/or merger.

Recent studies have demonstrated that gene regulatory rewiring often follows duplication (Gupta and Tsiantis, 2018), and that entire modules of genic coexpression may be duplicated and retained in plants. Ruprecht et al. (2016), for example, showed that in *Arabidopsis* a module for cell wall biosynthesis has become replicated and deployed differentially for different types of cell walls. Pfeifer et al. (2014) showed that genic coexpression in bread wheat grains was partitioned into 25 "modules", 23 of which contained biased suites of homoeologs from each of the hexaploid's three co-resident (A, B, D) genomes. More recently, Takahagi et al. (2018) conducted coexpression analysis of 727 RNAseq data sets from bread wheat, reiterating and extending these findings of differential homoeolog composition of key modules (here meaning suites of coexpressed genes) involved in many biological processes, including chloroplast biogenesis, RNA metabolism, putative defense response, putative posttranscriptional modification, and lipid metabolism. In contrast, Alabdullah et al. (2019) examined how polyploidization in wheat affected meiotic genes, using 130 RNA-seq samples to define co-expressed gene modules. Among the three modules significantly correlated with meiosis, most genes retained all three homoeologous copies, and genes within these modules also exhibited balanced homoeologous expression. Though the foregoing studies were not designed to evaluate whether homoeologous-genome modular portioning of genic coxpression arose prior to or subsequent to polyploidization, it is likely that these modular structures represent network level transcriptomic adjustment to the polyploid condition.

Explicit tests of this diploid-polyploid temporal partitioning have recently been performed using allopolyploid (AD-genome) cotton (*Gossypium*), which contains two genomes (A, D) which diverged 5–10 million years ago (mya) in diploid lineages that became reunited during polyploid formation 1–2 mya. Hu et al. (2016) studied gene coexpression networks in developing seeds of diploid as well as allopolyploid species, Network comparisons among species indicated that the global network topology of allopolyploid cotton was asymmetric, resembling one of its two diploid progenitors (the A-genome diploid) more than that of the other (D-genome parent). A novel feature of this study was that it extended, by example, concepts of homoeolog

bias, dominance, and transgressive expression to the network modular level. They further showed that the transcriptomic architecture in developing polyploid cotton seeds is a partial combination of the modules observed in the two diploid progenitors, and that domestication of wild allopolyploid cotton led to a more tightly integrated (more highly coexpressed) modular structure. These results for cotton seed development were recently extended to fiber development (Gallagher et al., 2020). Key results include the fact that notwithstanding a general preservation of network modular structure among the A- and D-genome diploids and the allopolyploid, fewer than a quarter of all homoeologs co-occured in the same module, showing substantial homoeologous expression rewiring (alteration of coexpression relationships) at the intramodular level. In addition, most modules exhibit D-homoeolog expression bias, with few showing A-homoeolog bias.

The preceding examples are illustrative of the various types of expression change that accompany polyploid formation, showing that not only is gene expression itself massively altered, but that this gene-level view has multiple parallels at the modular level of gene coexpression relationships. Yet, connections between these phenomena and demonstrations of adaptation or diversification remain mostly obscure, however, notwithstanding our growing understanding the relationships between modular genic content and biological processes. In this respect one promising approach is to combine coexpression network analysis with standard tools from population genetics, as exemplified in a marvelous recent study on *Theobroma cacao* from Brazil (Hämälä et al., 2020). Starting with the initial suggestion that evolutionary change often arises from allele frequency shifts simultaneously at multiple genes, Hämälä et al. (2020) studied genic coexpression relationships for 31 individuals from four geographically allopatric populations. Genes from modules enriched for specific biological processes were combined to explore whether they exhibited possible differential selection between populations, using a coexpression-module-based form of the widely used FST and dXY. They showed that modules associated with biological processes such as protein modification, flowering, and water transport were implicated in polygenic adaptation, "even though individual genes that are members of those groups do not bear strong signatures of selection." Noting that this example is for possible differential adaptation of populations of a single diploid species rather than for the effects of introgression or polyploidization, it conceptually helps point the way to identifying cases of adaptation stemming from the latter speciation and diversification processes.

CONCLUDING REMARKS

Decades of inquiry have generated much insight into the evolution of merged genomes following polyploidy and hybridization. For instance, we know that duplicated genes experience a diversity of evolutionary trajectories, being either lost, epigenetically silenced, or retained but with altered expression regulation and possible neofunctionalization or subfunctionalization, and that some of these outcomes may be tissue or organ-specific. For these and other consequences of genome merger discussed here, some have considered whether there might be evolutionary "rules" or mechanisms that are predictive of specific outcomes (Adams and Wendel, 2005; Doyle et al., 2008; Soltis et al., 2016; Wendel et al., 2018). One generality is that the early stages of genome merger and doubling profoundly impact the molecular, genomic and physiological machinery, as Barbara McClintock famously anticipated in her 1983 Nobel lecture, and as Feldman et al. (2012) captured in partitioning evolutionary change into "revolutionary," that is, arising shortly after genome merger, versus "evolutionary," namely change that accrues more gradually over time. Many empirical examples of phenotypic and genomic innovation have been discussed here, reflecting a broad spectrum of underlying mechanisms and phenotypes. Yet, notwithstanding the extraordinary advances of the last several decades and the increasing use of breathtaking technologies for probing genomes and transcriptomes, we still have only a rudimentary understanding of how genome merger generates phenotypic diversity and thereby contributes to evolutionary diversification. Also poorly understood is the relative importance of the many genomic responses discussed here to adaptive evolution or phenotypic innovation at the diploid vs. polyploid level. Both organismal processes entail genetic merger, which variously sets in motion a plethora of "omics" changes, as illustrated in **Figure 1**, but it is unknown which if any of these responses are more characteristic of diploid vs. polyploid evolution. Some progress in this direction has emerged from studies designed to assess this temporally partitioning, across multiple genera (e.g., Flagel et al., 2008; Xu et al., 2014; Edger et al., 2017; Zhang D. et al., 2018), but this clearly is an area that warrants further investigation.

It is of interest to consider the constraints that hinder our understanding of the genetic basis of phenotypic innovation that follows polyploidy. To be sure, we do not yet understand the full dimensionality of the genotype-to-phenotype equation, and thus our insight into phenotypes is limited by our present understanding of the propagation of information from the genome through all of the "omics" into something emergent that we call the phenotype (Casacuberta et al., 2016). In addition, though, we suggest that the early responses to genome merger and/or doubling represent only the tip of the iceberg compared to later evolutionary innovation, which ultimately was enabled or set in motion by genome merger but which remains latent, perhaps for millions of years, until ecological opportunity dovetails with novel genomic/omic recombinants. This temporal dimension is key to our perspective; genome merger sets the stage for both immediate and long-term evolutionary innovation. The retention of many to most duplicated genes and other genomic components in a polyploid serves as a massive reservoir of variation that may remain evolutionarily latent, perhaps for hundreds to thousands to millions of years. Only later, perhaps when exposed to altered selection pressures in novel environment, will this variation lead to phenotypic innovation, adaptation, and speciation. Yet this novel diversity may not have been possible without the ancient genomic infusion (or infusions) from interspecific gene flow. That is, long-term

retention of duplicated genes, regulatory elements, and other genomic components, may be responsible for evolutionary diversification, even after extinction of the parental lineages and in novel environments relative to the progenitors.

This possibility is consistent with the so-called radiation lag-time model following WGD, formulated on the basis of phylogenetic and divergence-time data (Schranz et al., 2012). According to this hypothesis, successful diversifications in groups that experience ancient polyploidy do not arise due to the sudden genesis of novel key traits, but instead reflects subsequent phenomena over evolutionary time, including changing environmental conditions. This perspective is also in line with classical views that genome doubling provides a massive reservoir of duplicated genes for longer-term evolution of new functions (Stebbins, 1950; Ohno, 1970). Finally, as we note above, the possibility of epigenetic "subfunctionalization" offers a mechanism for selective retention of duplicate genes for later release from suppression and evaluation by natural selection (Adams et al., 2003, 2004).

Thus, there is now a confluence among classical notions and modern genomic perspectives regarding the importance of long-term persistence of latent variation generated by hybridization and polyploidy on adaptation and diversification. Additionally, population genetic considerations are relevant, especially the extremely reduced effective population sizes that are involved in the early stages of polyploid formation and stabilization in many groups. These conditions minimize the importance of selection relative to drift, thus further facilitating the survival of less than perfectly adapted genomes and genotypes while highlighting the role of "chance" or stochasticity (Lynch, 2007) in generating genome complexity as well as biological diversity.

Predicting long-term effects of phenomena that have profound impact on organisms, and whose evolutionary fate is heavily dependent on spatio-temporal contexts, is not yet possible. Even retrospectively constructing the detailed history of these phenomena in current lineages is complicated and this has fueled conflicting views (Mayrose et al., 2011; Abbott et al., 2013; Butlin and Ritchie, 2013; Soltis P. S. et al., 2014). The evolutionary consequences of allopolyploidy and hybridization—particularly over the longer term—have been and will remain a matter of interest into the future. Yet we see promise in our growing understanding of biological processes ranging in scale from the molecular to the ecological. This enhanced understanding of the genotype-to-phenotype equation should increasingly inform comparative and ecological analyses of adaptation, thus permitting an improved appreciation of the temporal dynamics and genomic underpinnings of polyploidy-fueled diversification.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This study was supported by the grants CGL2017-88500-P (AEI/FEDER, EU) to GN and AGL2016-78992-R to JC from the Spanish Ministry of Economy and Competitivity, as well as Severo Ochoa Program for Centers of Excellence in R&D 2016–2020 (SEV-2015-0533) to JC, and by grants from the National Science Foundation to JW.

ACKNOWLEDGMENTS

We thank the United States and Spanish Fulbright Commissions as well as the Real Jardín Botánico and Center for Research in Agricultural Genomics, which supported JW during a sabbatical stay in Spain.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evolution. Biol.* 26, 229–246. doi: 10.1111/j.1420-9101.2012.02599.x

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *J. Syst. Evol.* 55, 238–258. doi: 10.1111/jse.12267

Abbott, R. J., Barton, N. H., and Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Mol. Ecol.* 25, 2325–2332. doi: 10.1111/mec.13685

Abbott, R. J., Hegarty, M. J., Hiscock, S. J., and Brennan, A. C. (2010). Homoploid hybrid speciation in action. *Taxon* 59, 1375–1386. doi: 10.1002/tax.595005

Adams, K. L., Cronn, R. C., Percifield, R., and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4649–4654. doi: 10.1073/pnas.0630618100

Adams, K. L., Percifield, R., and Wendel, J. F. (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168, 2217–2226. doi: 10.1534/genetics.104.033522

Adams, K. L., and Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8, 135–141. doi: 10.1016/j.pbi.2005.01.001

Aeschbacher, S., Selby, J. P., Willis, J. H., and Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7061–7066. doi: 10.1073/pnas.1616755114

Ågren, J. A., Huang, H. R., and Wright, S. I. (2016). Transposable element evolution in the allotetraploid capsella bursa-pastoris. *Am. J. Bot.* 103, 1197–1202. doi: 10.3732/ajb.1600103

Alabdullah, A. K., Borrill, P., Martin, A. C., Ramirez-Gonzalez, R. H., Hassani-Pak, K., Uauy, C., et al. (2019). A co-expression network in hexaploid wheat reveals mostly balanced expression and lack of significant gene loss of homeologous meiotic genes upon polyploidization. *Front. Plant Sci.* 10:1325. doi: 10.3389/fpls.2019.01325

Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I. J., Koblížková, A., et al. (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of Fritillaria lilies. *Ann. Bot.-London* 107, 255–268. doi: 10.1093/aob/mcq235

Arkhipova, I. R. (2018). Neutral theory, transposable elements, and eukaryotic genome evolution. *Mol. Biol. Evol.* 35, 1332–1337. doi: 10.1093/molbev/msy083

Arnold, M. L. (1993). *Iris nelsonii* (Iridaceae): origin and genetic composition of a homoploid hybrid species. *Am. J. Bot.* 80, 577–583. doi: 10.1002/j.1537-2197.1993.tb13843.x

Arnold, M. L. (1997). *Natural Hybridization and Evolution*. Oxford: Oxford University Press.

Arnold, M. L., Tang, S., Knapp, S. J., and Martin, N. H. (2010). Asymmetric introgressive hybridization among Louisiana *Iris* species. *Genes-Basel* 1, 9–22. doi: 10.3390/genes1010009

Baack, E. J., and Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Curr. Opin. Genet. Dev.* 17, 513–518. doi: 10.1016/j.gde.2007.09.001

Baduel, P., Quadrana, L., Hunter, B., Bomblies, K., and Colot, V. (2019). Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat. Commun.* 10:5818. doi: 10.1038/s41467-019-13730-0

Bao, Y., Hu, G., Grover, C. E., Conover, J., Yuan, D., and Wendel, J. F. (2019). Unraveling *cis* and *trans* regulatory evolution during cotton domestication. *Nat. Commun.* 10:5399. doi: 10.1038/s41467-019-13386-w

Bardil, A., Tayalé, A., and Parisod, C. (2015). Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex. *Plant J.* 82, 621–631. doi: 10.1111/tpj.12837

Barrier, M., Baldwin, B. G., Robichaux, R. H., and Purugganan, M. D. (1999). Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Biol. Evol.* 16, 1105–1113. doi: 10.1093/oxfordjournals.molbev.a026200

Barton, N. H. (2013). Does hybridization influence speciation? *J. Evolution. Biol.* 26, 267–269. doi: 10.1111/jeb.12015

Barton, N. H., and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16, 113–148.

Bay, R. A., and Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proc. R. Soc. B.* 284:20162414. doi: 10.1098/rspb.2016.2414

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., et al. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* 51, 877–884. doi: 10.1038/s41588-019-0405-z

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A. (2010). Heterosis. *Plant Cell* 22, 2105–2112. doi: 10.1105/tpc.110.076133

Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Blanco-Pastor, J. L., Vargas, P., and Pfeil, B. E. (2012). Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *PLoS One* 7:e39089. doi: 10.1371/journal.pone.0039089

Bock, D. G., Andrew, R. L., and Rieseberg, L. H. (2014). On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* 23, 4899–4911. doi: 10.1111/mec.12920

Bomblies, K., and Weigel, D. (2007). Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet.* 8, 382–393. doi: 10.1038/nrg2082

Bottani, S., Zabet, N. R., Wendel, J. F., and Veitia, R. A. (2018). Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* 23, 393–402. doi: 10.1016/j.tplants.2018.01.002

Brower, A. V. (2013). Introgression of wing pattern alleles and speciation via homoploid hybridization in *Heliconius* butterflies: a review of evidence from the genome. *Proc. R. Soc. B* 280:20122302. doi: 10.1098/rspb.2012.2302

Butlin, R. K., and Ritchie, M. G. (2013). Pulling together or pulling apart: hybridization in theory and practice. *J. Evol. Biol.* 26, 294–298. doi: 10.1111/jeb.12080

Carpentier, M.-C., Manfroi, E., Wei, F.-J., Wu, H.-P., Lasserre, E., Llauro, C., et al. (2019). Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat. Commun.* 10:24. doi: 10.1038/s41467-018-07974-5

Casacuberta, J. M., Jackson, S., Panaud, O., Purugganan, M., and Wendel, J. F. (2016). Evolution of plant phenotypes, from genomes to traits. *G3* 6, 775–778. doi: 10.1534/g3.115.025502

Casacuberta, J. M., and Santiago, N. (2003). Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311, 1–11. doi: 10.1016/S0378-1119(03)00557-2

Case, A. L., Finseth, F. R., Barr, C. M., and Fishman, L. (2016). Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proc. R. Soc. B* 283:20161493. doi: 10.1098/rspb.2016.1493

Cavé-Radet, A., Giraud, D., Lima, O., El Amrani, A., Aïnouche, M., and Salmon, A. (2019). Evolution of small RNA expression following hybridization and allopolyploidization: insights from *Spartina species* (Poaceae, Chloridoideae. *Plant Mol. Biol.* 102, 55–72. doi: 10.1007/s11103-019-00931-w

Chen, S., Ren, F., Zhang, L., Liu, Y., Chen, X., Li, Y., et al. (2018). Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. *Mol. Plant* 11, 914–927. doi: 10.1016/j.molp.2018.04.009

Chen, Z., Zhao, N., Li, S., Grover, C. E., Nie, H., Wendel, J. F., et al. (2017). Plant mitochondrial genome evolution and cytoplasmic male sterility. *Crit. Rev. Plant Sci.* 36, 55–69. doi: 10.1080/07352689.2017.1327762

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4, 258–268. doi: 10.1038/s41477-018-0136-7

Chester, M., Gallagher, J. P., Vaughan Symonds, V., Cruz da Silva, A. V., Mavrodiev, E. V., Leitch, A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. U.S.A.* 109, 1176–1181. doi: 10.1073/pnas.1112041109

Choudhury, R. R., Rogivue, A., Gugerli, F., and Parisod, C. (2019). Impact of polymorphic transposable elements on linkage disequilibrium along chromosomes. *Mol. Ecol.* 28, 1550–1562. doi: 10.1111/mec.15014

Christe, C., Stölting, K. N., Bresadola, L., Fussi, B., Heinze, B., Wegmann, D., et al. (2016). Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Mol. Ecol.* 25, 2482–2498. doi: 10.1111/mec.13587

Chuong, E. B., Elde, N. C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86. doi: 10.1038/nrg.2016.139

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Contreras-Lopez, O., Moyano, T. C., Soto, D. C., and Gutiérrez, R. A. (2018). "Step-by-step construction of gene co-expression networks from high-throughput arabidopsis RNA sequencing data," in *Root Development*, eds D. Ristova and E. Barbez (New York, NY: Humana Press), 275–301.

Corneillie, S., De Storme, N., Van Acker, R., Fangel, J. U., Bruyne, M., De Rycke, R., et al. (2019). Polyploidy affects plant growth and alters cell wall composition. *Plant Physiol.* 179, 74–87. doi: 10.1104/pp.18.00967

Corriveau, J. L., and Coleman, A. W. (1988). Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* 75, 1443–1458. doi: 10.1002/j.1537-2197.1988.tb11219.x

Crisp, P. A., Hammond, R., Zhou, P., Vaillancourt, B., Lipzen, A., Daum, C., et al. (2020). Variation and inheritance of small RNAs in maize inbreds and F1 hybrids. *Plant Physiol.* 182, 318–331. doi: 10.1104/pp.19.00817

Danilevicz, M. F., Fernandez, C. G. T., Marsh, J. I., Bayer, P. E., and Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* 54, 18–25. doi: 10.1016/j.pbi.2019.12.005

D'Ario, M., Griffiths-Jones, S., and Kim, M. (2017). Small RNAs: big impact on plant development. *Trends Plant Sci.* 22, 1056–1068. doi: 10.1016/j.tplants.2017.09.009

De Smet, R., Adams, K. L., Vandepoele, K., Van Montagu, M. C., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2898–2903. doi: 10.1073/pnas.1300127110

De Smet, R., Sabaghian, E., Li, Z., Saeys, Y., and Van de Peer, Y. (2017). Coordinated functional divergence of genes after genome duplication in *Arabidopsis thaliana*. *Plant Cell* 29, 2786–2800. doi: 10.1105/tpc.17.00531

Dickinson, H., Costa, L., and Gutierrez-Marcos, J. (2012). Epigenetic neofunctionalisation and regulatory gene evolution in grasses. *Trends Plant Sci.* 17, 389–394. doi: 10.1016/j.tplants.2012.04.002

Ding, M., and Chen, Z. J. (2018). Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Curr. Opin. Plant Biol.* 42, 37–48. doi: 10.1016/j.pbi.2018.02.003

Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* 35, 57–65. doi: 10.1016/j.gde.2015.09.001

Doyle, J. J., and Coate, J. E. (2019). Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* 180, 1–52. doi: 10.1086/700636

Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science* 366, 594–599. doi: 10.1126/science.aaw2090

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Ehrenreich, I. M., and Pfennig, D. W. (2016). Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Ann. Bot. Lond.* 117, 769–779. doi: 10.1093/aob/mcv130

El Baidouri, M., Do Kim, K., Abernathy, B., Arikit, S., Maumus, F., Panaud, O., et al. (2015). A new approach for annotation of transposable elements using small RNA mapping. *Nucleic Acids Res.* 43:e84. doi: 10.1093/nar/gkv257

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., et al. (2017). The genomic mosaicism of hybrid speciation. *Sci. Adv.* 3:e1602996. doi: 10.1126/sciadv.1602996

Ellstrand, N. C. (2014). Is gene flow the most important evolutionary force in plants? *Am. J. Bot.* 101, 737–753. doi: 10.3732/ajb.1400024

Emamjomeh, A., Robat, E. S., Zahiri, J., Solouki, M., and Khosravi, P. (2017). Gene co-expression network reconstruction: a review on computational methods for inferring functional information from plant-based expression data. *Plant Biotechnol. Rep.* 11, 71–86. doi: 10.1007/s11816-017-0433-z

Esnault, C., Lee, M., Ham, C., and Levin, H. L. (2019). Transposable element insertions in fission yeast drive adaptation to environmental stress. *Genome Res.* 29, 85–95. doi: 10.1101/gr.239699.118

Feldman, M., Levy, A., Chalhoub, B., and Kashkush, K. (2012). "Genomic plasticity in polyploid wheat," in *Polyploidy and Genome Evolution*, eds P. S. Soltis and D. E. Soltis (Heidelberg: Springer), 109–136.

Fishman, L., and Sweigart, A. L. (2018). When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annu. Rev. Plant Biol.* 69, 707–731. doi: 10.1146/annurev-arplant-042817-040113

Flagel, L. E., Udall, J., Nettleton, D., and Wendel, J. F. (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, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Flagel, L. E., Willis, J. H., and Vision, T. J. (2014). The standing pool of genomic structural variation in a natural population of *Mimulus guttatus*. *Genome Biol. Evol.* 6, 53–64. doi: 10.1093/gbe/evt199

Flood, P., Theeuwen, T., Schneeberger, K., Keizer, P., Severing, E., Kouklas, E., et al. (2020). Reciprocal cybrids reveal how organellar genomes affect plant phenotypes Authors. *Nat. Plants* 6, 13–21. doi: 10.1038/s41477-019-0575-9

Folk, R. A., Soltis, P. S., Soltis, D. E., and Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *Am. J. Bot.* 105, 364–375. doi: 10.1002/ajb2.1018

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., et al. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347:1258524. doi: 10.1126/science.1258524

Fuentes, R. R., Chebotarov, D., Duitama, J., Smith, S., De la Hoz, J. F., Mohiyuddin, M., et al. (2019). Structural variants in 3000 rice genomes. *Genome Res.* 29, 870–880. doi: 10.1101/gr.241240.118

Gabur, I., Chawla, H. S., Snowdon, R. J., and Parkin, I. A. (2019). Connecting genome structural variation with complex traits in crop plants. *Theor. Appl. Genet.* 132, 733–750. doi: 10.1007/s00122-018-3233-0

Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., and Osborn, T. C. (2007). Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19, 3403–3417. doi: 10.1105/tpc.107.054346

Galindo-González, L., Mhiri, C., Deyholos, M. K., and Grandbastien, M. A. (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene* 160, 67–75. doi: 10.1016/j.gene.2017.04.051

Gallagher, J. P., Hu, G., Grover, C. E., Jareczek, J. J., and Wendel, J. F. (2020). Effects of genome duplication and domestication on the fiber gene coexpression network of *Gossypium hirsutum* L. *G3* 10:1362.

García, N., Folk, R. A., Meerow, A. W., Chamala, S., Gitzendanner, M. A., de Oliveira, R. S., et al. (2017). Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Mol. Phylogenet. Evol.* 111, 231–247. doi: 10.1016/j.ympev.2017.04.003

Göbel, U., Arce, A. L., He, F., Rico, A., Schmitz, G., and De Meaux, J. (2018). Robustness of transposable element regulation but no genomic shock observed in interspecific *Arabidopsis* hybrids. *Genome Biol. Evol.* 10, 1403–1415. doi: 10.1093/gbe/evy095

Golicz, A. A., Batley, J., and Edwards, D. (2016). Towards plant pangenomics. *Plant Biotechnol. J.* 14, 1099–1105. doi: 10.1111/pbi.12499

Gong, L., Olson, M., and Wendel, J. F. (2014). Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol. Biol. Evol.* 31, 2624–2636. doi: 10.1093/molbev/msu207

Gong, L., Salmon, A., Yoo, M. J., Grupp, K. K., Wang, Z., Paterson, A. H., et al. (2012). The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. *Mol. Biol. Evol.* 29, 3023–3036. doi: 10.1093/molbev/mss110

Gottlieb, L. D. (1984). Genetics and morphological evolution in plants. *Am. Nat.* 123, 681–709. doi: 10.1086/284231

Gout, J. F., and Lynch, M. (2015). Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol. Biol. Evol.* 32, 2141–2148. doi: 10.1093/molbev/msv095

Grant, B. R., and Grant, P. R. (2017). Watching speciation in action. *Science* 355, 910–911. doi: 10.1126/science.aam6411

Grant, P. R., and Grant, B. R. (1994). Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution* 48, 297–316. doi: 10.1111/j.1558-5646.1994.tb01313.x

Grant, P. R., and Grant, B. R. (2009). The secondary contact phase of allopatric speciation in Darwin's finches. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20141–20148. doi: 10.1073/pnas.0911761106

Grant, V. (1981). *Plant speciation*. New York, NY: Columbia University Press.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722. doi: 10.1126/science.1188021

Grob, S. (2020). Three-dimensional chromosome organization in flowering plants. *Brief. Funct. Genomics* 19, 83–91. doi: 10.1093/bfgp/elz024

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Guan, X., Pang, M., Nah, G., Shi, X., Ye, W., Stelly, D. M., et al. (2014). miR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat. Commun.* 5:3050. doi: 10.1038/ncomms4050

Gupta, M. D., and Tsiantis, M. (2018). Gene networks and the evolution of plant morphology. *Curr. Opin. Plant Biol.* 45, 82–87. doi: 10.1016/j.pbi.2018.05.011

Gutzat, R., and Mittelsten Scheid, O. (2012). Epigenetic responses to stress: triple defense? *Curr. Opin. Plant Biol.* 15, 568–573. doi: 10.1016/j.pbi.2012.08.007

Hämälä, T., Guiltinan, M. J., Marden, J. H., Maximova, S. N., dePamphilis, C. W., and Tiffin, P. (2020). Gene expression modularity reveals footprints of polygenic adaptation in *Theobroma cacao*. *Mol. Biol. Evol.* 37, 110–123. doi: 10.1093/molbev/msz206

Hamilton, J. A., and Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv. Biol.* 30, 33–41. doi: 10.1111/cobi.12574

Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford Surv. Evol. Biol.* 7, 69–128.

Harrison, R. G. (ed.) (1993). *Hybrid Zones and the Evolutionary Process*. New York, NY: Oxford University Press.

Hawkins, J. S., Proulx, S. R., Rapp, R. A., and Wendel, J. F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon

proliferation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17811–17816. doi: 10.1073/pnas.0904339106

He, L., Wu, W., Zinta, G., Yang, L., Wang, D., Liu, R., et al. (2018). A naturally occurring epiallele associates with leaf senescence and local climate adaptation in *Arabidopsis accessions*. *Nat. Commun.* 9:460. doi: 10.1038/s41467-018-02839-3

Heliconius Genome, and Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98. doi: 10.1038/nature11041

Hochholdinger, F., and Baldauf, J. A. (2018). Heterosis in plants. *Curr. Biol.* 28, R1089–R1092. doi: 10.1016/j.cub.2018.06.041

Hodges, S. A., Burke, J. M., and Arnold, M. L. (1996). Natural formation of *Iris* hybrids: experimental evidence on the establishment of hybrid zones. *Evolution* 50, 2504–2509. doi: 10.2307/2410717

Hof, A. E. V. T., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., et al. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105. doi: 10.1038/nature17951

Hohmann, N., and Koch, M. A. (2017). An *Arabidopsis* introgression zone studied at high spatio-temporal resolution: interglacial and multiple genetic contact exemplified using whole nuclear and plastid genomes. *BMC Genomics* 18:810. doi: 10.1186/s12864-017-4220-6

Hopkins, R. (2013). Reinforcement in plants. *New Phytol.* 197, 1095–1103. doi: 10.1111/nph.12119

Hu, G., Hovav, R., Grover, C. E., Faigenboim-Doron, A., Kadmon, N., Page, J. T., et al. (2016). Evolutionary conservation and divergence of gene coexpression networks in *Gossypium* (cotton) seeds. *Genome Biol. Evol.* 8, 3765–3783. doi: 10.1093/gbe/evw280

Hu, G., and Wendel, J. F. (2019). Cis–trans controls and regulatory novelty accompanying allopolyploidization. *New Phytol.* 221, 1691–1700. doi: 10.1111/nph.15515

Huang, C., Sun, H., Xu, D., Chen, Q., Liang, Y., Wang, X., et al. (2017). ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl. Acad. Sci. U.S.A.* 115, E334–E341. doi: 10.1073/pnas.1718058115

Huang, K., and Rieseberg, L. H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Front. Plant Sci.* 11:296. doi: 10.3389/fpls.2020.00296

Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., et al. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62. doi: 10.1038/s41477-018-0329-0

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C. A., Carretero-Paulet, L., Chang, T.-H., et al. (2013). Architecture and evolution of a minute plant genome. *Nature* 498, 94–98. doi: 10.1038/nature12132

Janzen, G. M., Wang, L., and Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytol.* 221, 1279–1288. doi: 10.1111/nph.15457

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi: 10.1038/nature21370

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Joehanes, R. (2018). "Network analysis of gene expression," in *N. Raghavachari and N. Garcia-Reyero*, eds Gene Expression and Analysis (New York, NY: Humana Press), 325–341.

Joly, S., McLenachan, P. A., and Lockhart, P. J. (2009). A statistical approach for distinguishing hybridization and incomplete lineage sorting. *Am. Nat.* 174, E54–E70. doi: 10.1086/600082

Jouffroy, O., Saha, S., Mueller, L., Quesneville, H., Maumus, F., Maumus, F., et al. (2016). Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC Genomics* 17:624. doi: 10.1186/s12864-016-2980-z

Kashkush, K., Feldman, M., and Levy, A. A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* 33, 102–106. doi: 10.1038/ng1063

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

Kim, M., Cui, M. L., Cubas, P., Gillies, A., Lee, K., Chapman, M. A., et al. (2008). Regulatory genes control a key morphological and ecological trait transferred between species. *Science* 322, 1116–1119. doi: 10.1126/science.1164371

Kim, M. Y., and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. *Trends Plant Sci.* 19, 320–326. doi: 10.1016/j.tplants.2014.01.014

Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18, 210.

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., and Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science* 359, 224–228. doi: 10.1126/science.aao4593

Lanciano, S., and Mirouze, M. (2018). Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Curr. Opin. Genet. Dev.* 49, 106–114. doi: 10.1016/j.gde.2018.04.002

Landis, J. B., Soltis, D. E., Zheng, L., Marx, H. E., Tank, D., Barker, M. S., et al. (2018). Impact of whole genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 433–444. doi: 10.1002/ajb2.1060

Leducq, J. B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J. P., Samani, P., et al. (2016). Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nat. Microbiol.* 1, 15003. doi: 10.1038/nmicrobiol.2015.3

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Lermontova, I., Sandmann, M., Mascher, M., Schmit, A. C., and Chabouté, M. E. (2015). Centromeric chromatin and its dynamics in plants. *Plant J.* 83, 4–17. doi: 10.1111/tpj.12875

Leroy, T., Louvet, J. M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J. M., et al. (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytol.* 226, 1171–1182. doi: 10.1111/nph.16095

Levin, D. A. (1983). Polyploidy and novelty in flowering plants. *Am. Nat.* 122, 1–25.

Levin, D. A., Francisco-Ortega, J., and Jansen, R. K. (1996). Hybridization and the extinction of rare plant species. *Conserv. Biol.* 10, 10–16. doi: 10.1046/j.1523-1739.1996.10010010.x

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., et al. (2014). mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* 26, 1878–1900. doi: 10.1105/tpc.114.124388

Li, C., Sun, X., Conover, J. L., Zhang, Z., Wang, J., Wang, X., et al. (2019). Cytonuclear coevolution following homoploid hybrid speciation in *Aegilops tauschii*. *Mol. Biol. Evol.* 36, 341–349. doi: 10.1093/molbev/msy215

Li, N., Xu, C., Zhang, A., Lv, R., Meng, X., Lin, X., et al. (2019). DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytol.* 223, 979–992. doi: 10.1111/nph.15820

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., and De Smet, R. (2016). Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* 28, 326–344. doi: 10.1105/tpc.15.00877

Liang, Y., Liu, Q., Wang, X., Huang, C., Xu, G., Hey, S., et al. (2019). ZmMADS69 functions as a flowering activator through the ZmRap2.7-ZCN8 regulatory module and contributes to maize flowering time adaptation. *New Phytol.* 221, 2335–2347. doi: 10.1111/nph.15512

Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. doi: 10.1038/nrg3374

Liu, Q., Deng, S., Liu, B., Tao, Y., Ai, H., Liu, J., et al. (2020). A helitron-induced RabGDIα variant causes quantitative recessive resistance to maize rough dwarf disease. *Nat. Commun.* 11:495. doi: 10.1038/s41467-020-14372-3

Liu, S. L., and Adams, K. L. (2010). Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. *Mol. Biol. Evol.* 27, 2817–2828. doi: 10.1093/molbev/msq169

Lohaus, R., and Van de Peer, Y. (2016). Of dups and dinos: evolution at the K/Pg boundary. *Curr. Opin. Plant Biol.* 30, 62–69. doi: 10.1016/j.pbi.2016.01.006

Lopes, F. R., Jjingo, D., Da Silva, C. R. M., Andrade, A. C., Marraccini, P., Teixeira, J. B., et al. (2013). Transcriptional activity, chromosomal distribution and expression effects of transposable elements in Coffea genomes. *PLoS One* 7:e35143. doi: 10.1371/journal.pone.0078931

Lowry, D. B., and Gould, B. A. (2016). "Speciation continuum," in *Encyclopedia of Evolutionary Biology*, ed. R. Kliman (New York, NY: Academic Press), 159–165.

Lu, H., Xue, L., Cheng, J., Yang, X., Xie, H., Song, X., et al. (2020). Polyploidization-driven differentiation of freezing tolerance in *Solidago canadensis*. *Plant Cell Environ.* 43, 1394–1403. doi: 10.1111/pce.13745

Luo, D., Carpenter, R., Vincent, C., Copsey, L., and Coen, E. (1996). Origin of floral asymmetry in *Antirrhinum*. *Nature* 383, 794–799. doi: 10.1038/383794a0

Lv, Y., Hu, F., Zhou, Y., Wu, F., and Gaut, B. S. (2019). Maize transposable elements contribute to long non-coding RNAs that are regulatory hubs for abiotic stress response. *BMC Genomics* 20:864. doi: 10.1186/s12864-019-6245-5

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8597–8604. doi: 10.1073/pnas.0702207104

Ma, Y., Wang, J., Hu, Q., Li, J., Sun, Y., Zhang, L., et al. (2019). Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Commun. Biol.* 2:213. doi: 10.1038/s42003-019-0445-z

Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., et al. (2005). Genomic changes in synthetic Arabidopsis polyploids. *Plant J.* 41, 221–230. doi: 10.1111/j.1365-313X.2004.02297.x

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi: 10.1038/nature05706

Marburger, S., Monnahan, P., Seear, P. J., Martin, S. H., Koch, J., Paajanen, P., et al. (2019). Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* 10:5218. doi: 10.1038/s41467-019-13159-5

Marques, I., Fuertes Aguilar, J., Martins-Louçao, M. A., Moharrek, F., and Nieto Feliner, G. (2017). A three-genome five-gene comprehensive phylogeny of the bulbous genus *Narcissus* (Amaryllidaceae) challenges current classifications and reveals multiple hybridization events. *Taxon* 66, 832–854. doi: 10.12705/664.3

Martin, N., Bouck, A., and Arnold, M. (2006). Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* 172, 2481–2489. doi: 10.1534/genetics.105.053538

Martin, S. H., and Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* 47, 69–74. doi: 10.1016/j.gde.2017.08.007

Martínez, G., and Slotkin, R. K. (2012). Developmental relaxation of transposable element silencing in plants: Functional or byproduct? *Curr. Opin. Plant Biol.* 15, 496–502. doi: 10.1016/j.pbi.2012.09.001

Mavárez, J., and Linares, M. (2008). Homoploid hybrid speciation in animals. *Mol. Ecol.* 17, 4181–4185. doi: 10.1111/j.1365-294X.2008.03898.x

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: The Belknap Press of Harvard University Press.

Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., et al. (2011). Recently formed polyploid plants diversify at lower rates. *Science* 333, 1257–1257. doi: 10.1126/science.1207205

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

McCue, A. D., and Slotkin, R. K. (2012). Transposable element small RNAs as regulators of gene expression. *Trends Genet.* 28, 616–623. doi: 10.1016/j.tig.2012.09.001

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363. doi: 10.1038/ncomms14363

Mhiri, C., Parisod, C., Daniel, J., Petit, M., Lim, K. Y., Dorlhac de Borne, F., et al. (2019). Parental transposable element loads influence their dynamics in young *Nicotiana* hybrids and allotetraploids. *New Phytol.* 221, 1619–1633. doi: 10.1111/nph.15484

Mitchell, N., Owens, G. L., Hovick, S. M., Rieseberg, L. H., and Whitney, K. D. (2019). Hybridization speeds adaptive evolution in an eight-year field experiment. *Sci. Rep.* 9:6746. doi: 10.1038/s41598-019-43119-4

Monnahan, P., Koláø, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., et al. (2019). Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* 3, 457–468. doi: 10.1038/s41559-019-0807-4

Morata, J., Marín, F., Payet, J., and Casacuberta, J. M. (2018a). Plant lineage-specific amplification of transcription factor binding motifs by miniature inverted-repeat transposable elements (MITEs). *Genome Biol. Evol.* 10, 1210–1220. doi: 10.1093/gbe/evy073

Morata, J., Tormo, M., Alexiou, K. G., Vives, C., Ramos-Onsins, S. E., Garcia-Mas, J., et al. (2018b). The evolutionary consequences of transposon-related pericentromere expansion in melon. *Genome Biol. Evol.* 10, 1584–1595. doi: 10.1093/gbe/evy115

Ng, D. W., Lu, J., and Chen, Z. J. (2012). Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr. Opin. Plant Biol.* 15, 154–161. doi: 10.1016/j.pbi.2012.01.007

Ni, Z., Kim, E. D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457, 327–331. doi: 10.1038/nature07523

Nieto Feliner, G., Álvarez, I., Fuertes Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., et al. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* 118, 513–516. doi: 10.1038/hdy.2017.7

Niu, X. M., Xu, Y. C., Li, Z. W., Bian, Y. T., Hou, X. H., Chen, J. F., et al. (2019). Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 6908–6913. doi: 10.1073/pnas.1811498116

Noshay, J. M., Anderson, S. N., Zhou, P., Ji, L., Ricci, W., Lu, Z., et al. (2019). Monitoring the interplay between transposable element families and DNA methylation in maize. *PLoS Genet.* 15:e1008291. doi: 10.1371/journal.pgen.1008291

Ohno, S. (1970). *Evolution by Gene Duplication*. New York, NY: Springer-Verlag.

Ottenburghs, J. (2018). Exploring the hybrid speciation continuum in birds. *Ecol. Evol.* 8, 13027–13034. doi: 10.1002/ece3.4558

Palacios, P. M., Jacquemot, M. P., Tapie, M., Rousselet, A., Diop, M., Remoue, C., et al. (2019). Assessing the response of small RNA populations to allopolyploidy using resynthesized *Brassica napus* allotetraploids. *Mol. Biol. Evol.* 36, 709–726. doi: 10.1093/molbev/msz007

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523

Panchy, N. L., Azodi, C. B., Winship, E. F., O'Malley, R. C., and Shiu, S. H. (2019). Expression and regulatory asymmetry of retained *Arabidopsis thaliana* transcription factor genes derived from whole genome duplication. *BMC Evol. Biol.* 19:77. doi: 10.1186/s12862-019-1398-z

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., et al. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Mol. Ecol.* 22, 3304–3317. doi: 10.1111/mec.12201

Payseur, B. A., and Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Mol. Ecol.* 25, 2337–2360. doi: 10.1111/mec.13557

Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14:e1002379. doi: 10.1371/journal.pbio.1002379

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes-Basel.* 9:88. doi: 10.3390/genes9020088

Pennisi, E. (2016). A shortcut to a species. *Science* 354:818. doi: 10.1126/science.354.6314.818

Pereira, J. F., and Ryan, P. R. (2019). The role of transposable elements in the evolution of aluminium resistance in plants. *J. Exp. Bot.* 70, 41–54. doi: 10.1093/jxb/ery357

Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q. T., et al. (2010). Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytol.* 186, 135–147. doi: 10.1111/j.1469-8137.2009.03140.x

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., et al. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345:1250091. doi: 10.1126/science.1250091

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., et al. (2006). Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* 16, 1262–1269. doi: 10.1101/gr.5290206

Poretti, M., Praz, C. R., Meile, L., Kälin, C., Schaefer, L. K., Schläfli, M., et al. (2020). Domestication of high-copy transposons underlays the wheat small RNA response to an obligate pathogen. *Mol. Biol. Evol.* 37, 839–848. doi: 10.1093/molbev/msz272

Quadrana, L., Etcheverry, M., Gilly, A., Caillieux, E., Madoui, M.-A., Guy, J., et al. (2019). Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nat. Commun.* 10:3421. doi: 10.1038/s41467-019-11385-5

Rao, X., and Dixon, R. A. (2019). Co-Expression networks for plant biology: why and how. *Acta Biochem. Biophys. Sin.* 51, 981–988. doi: 10.1093/abbs/gmz080

Rech, G. E., Bogaerts-Márquez, M., Barrón, M. G., Merenciano, M., Villanueva-Cañas, J. L., Horváth, V., et al. (2019). Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet.* 10:3421. doi: 10.1371/journal.pgen.1007900

Rensing, S. A. (2014). Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* 17, 43–48. doi: 10.1016/j.pbi.2013.11.002

Richter, T. E., and Ronald, P. C. (2000). The evolution of disease resistance genes. *Plant Mol. Biol.* 42, 195–204. doi: 10.1023/A:1006388223475

Rieseberg, L. H. (1991). Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *Am. J. Bot.* 78, 1218–1237. doi: 10.1002/j.1537-2197.1991.tb11415.x

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216. doi: 10.1126/science.1086949

Rieseberg, L. H., and Wendel, J. F. (1993). "Introgression and its consequences in plants," in *Hybrid Zones and the Evolutionary Process*, ed. R. Harrison (Oxford: Oxford University Press), 70–109.

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., and Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14:e2000234. doi: 10.1371/journal.pbio.2000234

Roux, F., Mary-Huard, T., Barillot, E., Wenes, E., Botran, L., Durand, S., et al. (2016). Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3687–3692. doi: 10.1073/pnas.1520687113

Runemark, A., Vallejo-Marin, M., and Meier, J. I. (2019). Eukaryote hybrid genomes. *PLoS Genet.* 15:e1008404. doi: 10.1371/journal.pgen.1008404

Ruprecht, C., Lohaus, R., Vanneste, K., Mutwil, M., Nikoloski, Z., Van de Peer, Y., et al. (2017a). Revisiting ancestral polyploidy in plants. *Sci. Adv.* 3:e1603195. doi: 10.1126/sciadv.1603195

Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S. R., Fernie, A. R., et al. (2016). FamNet: a framework to identify multiplied modules driving pathway expansion in plants. *Plant Physiol.* 170, 1878–1894. doi: 10.1104/pp.15.01281

Ruprecht, C., Vaid, N., Proost, S., Persson, S., and Mutwil, M. (2017b). Beyond genomics: studying evolution with gene coexpression networks. *Trends Plant Sci.* 22, 298–307. doi: 10.1016/j.tplants.2016.12.011

Salazar, C., Baxter, S. W., Pardo-Diaz, C., Wu, G., Surridge, A., Linares, M., et al. (2010). Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genet* 6:e1000930. doi: 10.1371/journal.pgen.1000930

Samans, B., Snowdon, R., and Mason, A. S. (2018). "Homoeologous exchanges and gene losses generate diversity and differentiate the B. napus genome from that of its ancestors," in *The Brassica napus* Genome, eds S. Liu, R. Snowdon, and B. Chalhoub (Cham: Springer), 131–148.

Sang, T., Crawford, D. J., and Stuessy, T. F. (1995). Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6813–6817. doi: 10.1073/pnas.92.15.6813

Sarilar, V., Palacios, P. M., Rousselet, A., Ridel, C., Falque, M., Eber, F., et al. (2013). Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. *New Phytol.* 198, 593–604. doi: 10.1111/nph.12156

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119

Sætre, G. P. (2013). Hybridization is important in evolution, but is speciation? *J. Evol. Biol.* 26, 256–258. doi: 10.1111/jeb.12005

Saxena, R. K., Edwards, D., and Varshney, R. K. (2014). Structural variations in plant genomes. *Brief. Funct. Genomics* 13, 296–307. doi: 10.1093/bfgp/elu016

Scascitelli, M., Whitney, K., Randell, R., King, M., Buerkle, C., and Rieseberg, L. H. (2010). Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Mol. Ecol.* 19, 521–541. doi: 10.1111/j.1365-294X.2009.04504.x

Schiessl, S. V., Katche, E., Ihien, E., Chawla, H. S., and Mason, A. S. (2019). The role of genomic structural variation in the genetic improvement of polyploid crops. *Crop J.* 7, 127–140. doi: 10.1016/j.cj.2018.07.006

Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., and Melzer, R. (2020). Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 225, 511–529. doi: 10.1111/nph.16122

Schmickl, R., Marburger, S., Bray, S., and Yant, L. (2017). Hybrids and horizontal transfer: introgression allows adaptive allele discovery. *J. Exp. Bot.* 68, 5453–5470. doi: 10.1093/jxb/erx297

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., et al. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Commun.* 5:5495. doi: 10.1038/ncomms6495

Schrader, L., and Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Mol. Ecol.* 28, 1537–1549. doi: 10.1111/mec.14794

Schranz, M. E., Mohammadin, S., and Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr. Opin. Plant Biol.* 15, 147–153. doi: 10.1016/j.pbi.2012.03.011

Schumer, M., Cui, R., Powell, D. L., Rosenthal, G. G., and Andolfatto, P. (2016). Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* 25, 2661–2679. doi: 10.1111/mec.13602

Schumer, M., Rosenthal, G. G., and Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution* 68, 1553–1560. doi: 10.1111/evo.12399

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360, 656–660. doi: 10.1126/science.aar3684

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., et al. (2014). Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176–192. doi: 10.1038/nrg3644

Sehrish, T., Symonds, V. V., Soltis, D. E., Soltis, P. S., and Tate, J. A. (2015). Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One* 10:e0144339. doi: 10.1371/journal.pone.0144339

Senerchia, N., Felber, F., and Parisod, C. (2014). Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytol.* 202, 975–985. doi: 10.1111/nph.12731

Serin, E. A., Nijveen, H., Hilhorst, H. W., and Ligterink, W. (2016). Learning from co-expression networks: possibilities and challenges. *Front. Plant Sci.* 7:444. doi: 10.3389/fpls.2016.00444

Servedio, M. R., Hermisson, J., and Van Doorn, G. S. (2013). Hybridization may rarely promote speciation. *J. Evol. Biol.* 26, 282–285. doi: 10.1111/jeb.12038

Sharbrough, J., Conover, J. L., Tate, J. A., Wendel, J. F., and Sloan, D. B. (2017). Cytonuclear responses to genome doubling. *Am. J. Bot.* 104, 1277–1280. doi: 10.3732/ajb.1700293

Sharma, B., and Kramer, E. (2013). Sub-and neo-functionalization of APETALA 3 paralogs have contributed to the evolution of novel floral organ identity in *Aquilegia* (columbine, Ranunculaceae). *New Phytol.* 197, 949–957. doi: 10.1111/nph.12078

Simonin, K. A., and Roddy, A. B. (2018). Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biol.* 16:e2003706. doi: 10.1371/journal.pbio.2003706

Soltis, D. E., Segovia-Salcedo, M. C., Jordon-Thaden, I., Majure, L., Miles, N. M., Mavrodiev, E. V., et al. (2014). Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol.* 202, 1105–1117. doi: 10.1111/nph.12756

Soltis, D. E., Visger, C. J., Marchant, D. B., and Soltis, P. S. (2016). Polyploidy: pitfalls and paths to a paradigm. *Am. J. Bot.* 103, 1146–1166. doi: 10.3732/ajb.1500501

Soltis, P. S. (2013). Hybridization, speciation and novelty. *J. Evol. Biol.* 26, 291–293. doi: 10.1111/jeb.12095

Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J., and Soltis, D. E. (2014). Polyploidy and novelty: Gottlieb's legacy. *Philos. T. R. Soc. B* 369, 20130351. doi: 10.1098/rstb.2013.0351

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Soltis, P. S., and Soltis, D. E. (2012). *Polyploidy and Genome Evolution*. Berlin: Springer.

Song, Q., and Chen, Z. J. (2015). Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* 24, 101–109. doi: 10.1016/j.pbi.2015.02.007

Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18, 99. doi: 10.1186/s13059-017-1229-8

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant–environment interactions. *Annu. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334

Springer, N. M., Li, Q., and Lisch, D. (2016). Creating order from chaos: epigenome dynamics in plants with complex genomes. *Plant Cell* 28, 314–325. doi: 10.1105/tpc.15.00911

Stebbins, G. L. (1940). The significance of polyploidy in plant evolution. *Am. Nat.* 74, 54–66. doi: 10.1086/280872

Stebbins, G. L. (1950). *Variation and Evolution in Plants*. New York, NY: Columbia.

Stebbins, G. L. (1971). *Chromosomal evolution in higher plants*. London: Edward Arnold.

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C., et al. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Mol. Ecol.* 25, 2427–2442. doi: 10.1111/mec.13539

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Douglas, C. J., and Cronk, Q. C. (2018a). Introgression from *Populus balsamifera* underlies adaptation and range boundaries in *P. trichocarpa*. *New Phytol.* 217, 416–427. doi: 10.1111/nph.14779

Suarez-Gonzalez, A., Lexer, C., and Cronk, Q. C. (2018b). Adaptive introgression: a plant perspective. *Biol. Lett.* 14:20170688. doi: 10.1098/rsbl.2017.0688

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

Swanson-Wagner, R. A., Eichten, S. R., Kumari, S., Tiffin, P., Stein, J. C., and Ware, D. (2010). Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* 20, 1689–1699. doi: 10.1101/gr.109165.110

Takahagi, K., Inoue, K., and Mochida, K. (2018). Gene co-expression network analysis suggests the existence of transcriptional modules containing a high proportion of transcriptionally differentiated homoeologs in hexaploid wheat. *Front. Plant Sci.* 9:1163. doi: 10.3389/fpls.2018.01163

Tan, S., Zhong, Y., Hou, H., Yang, S., and Tian, D. (2012). Variation of presence/absence genes among *Arabidopsis* populations. *BMC Evol. Biol.* 12:86. doi: 10.1186/1471-2148-12-86

Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., and Hinchliff, C. E. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467. doi: 10.1111/nph.13491

Twyford, A. D., and Ennos, R. A. (2012). Next-generation hybridization and introgression. *Heredity* 108, 179–189. doi: 10.1038/hdy.2011.68

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

Vanneste, K., Maere, S., and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the *Cretaceous* and the consequences for plant evolution. *Philos. T. R. Soc. B* 369:20130353. doi: 10.1098/rstb.2013.0353

Vicient, C. M., and Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* 120, 195–207. doi: 10.1093/aob/mcx078

Wagner, C. E. (2018). Improbable big birds. *Science* 359, 157–159. doi: 10.1126/science.aar4796

Wang, M., Wang, P., Lin, M., Ye, Z., Li, G., and Tu, L. (2018). Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* 4, 90–97. doi: 10.1038/s41477-017-0096-3

Wang, N., Wang, H., Wang, H., Zhang, D., Wu, Y., Ou, X., et al. (2010). Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with *Zizania latifolia*. *BMC Plant Biol.* 10:190. doi: 10.1186/1471-2229-10-190

Wang, X. R., Szmidt, A. E., and Savolainen, O. (2001). Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan Plateau. *Genetics* 159, 337–346.

Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., and Xu, J. (2015). Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* 47, 944–948. doi: 10.1038/ng.3346

Wei, Y., Li, F., Zhang, S., Zhang, S., Zhang, H., and Sun, R. (2019). Analysis of small RNA changes in different *Brassica napus* synthetic allopolyploids. *PeerJ* 7:e7621. doi: 10.7717/peerj.7621

Welch, M. E., and Rieseberg, L. H. (2002). Habitat divergence between a homoploid hybrid sunflower species, *Helianthus paradoxus* (Asteraceae), and its progenitors. *Am. J. Bot.* 89, 472–478. doi: 10.3732/ajb.89.3.472

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Wendel, J. F., Jackson, S. A., Meyers, B. C., and Wing, R. A. (2016). Evolution of plant genome architecture. *Genome Biol.* 17:37. doi: 10.1186/s13059-016-0908-1

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Whitney, K. D., Randell, R. A., and Rieseberg, L. H. (2010). Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol.* 187, 230–239. doi: 10.1111/j.1469-8137.2010.03234.x

Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., and Larson, T. R. (2012). A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noncapine. *Science* 336, 1704–1708. doi: 10.1126/science.1220757

Wu, M., Ge, Y., Xu, C., and Wang, J. (2020). Metabolome and transcriptome analysis of hexaploid *Solidago canadensis* roots reveals its invasive capacity related to polyploidy. *Genes Basel* 11:187. doi: 10.3390/genes11020187

Wu, S., Han, B., and Jiao, Y. (2019). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant.* 13, 59–71. doi: 10.1016/j.molp.2019.10.012

Xie, F., and Zhang, B. (2015). micro RNA evolution and expression analysis in polyploidized cotton genome. *Plant Biotechnol. J.* 13, 421–434. doi: 10.1111/pbi.12295

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

Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., et al. (2017). Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6133–6138. doi: 10.1073/pnas.1700073114

Yaakov, B., and Kashkush, K. (2012). Mobilization of Stowaway-like MITEs in newly formed allohexaploid wheat species. *Plant Mol. Biol.* 80, 419–427. doi: 10.1007/s11103-012-9957-3

Yakimowski, S. B., and Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *Am. J. Bot.* 101, 1247–1258. doi: 10.3732/ajb.1400201

Yoo, M. J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Yu, Y., Zhang, Y., Chen, X., and Chen, Y. (2019). Plant noncoding RNAs: hidden Players in development and stress responses. *Annu. Rev. Cell Dev. Bi.* 35, 407–431. doi: 10.1146/annurev-cellbio-100818-125218

Zhai, Y., Yu, X., Zhu, Z., Wang, P., Meng, Y., Zhao, Q., et al. (2019). Nuclear–cytoplasmic coevolution analysis of RuBisCO in synthesized *Cucumis* allopolyploid. *Genes-Basel* 10:869. doi: 10.3390/genes10110869

Zhang, D., Pan, Q., Tan, C., Liu, L., Ge, X., Li, Z., et al. (2018). Homoeolog expression is modulated differently by different subgenomes in *Brassica napus* hybrids and allotetraploids. *Plant Mol. Biol. Rep.* 36, 387–398. doi: 10.1007/s11105-018-1087-x

Zhang, H., Lang, Z., and Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z

Zhang, H., Zheng, R., Wang, Y., Zhang, Y., Hong, P., Fang, Y., et al. (2019). The effects of *Arabidopsis* genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Res.* 47, 7857–7869. doi: 10.1093/nar/gkz511

Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J., Huneau, C., et al. (2011). Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18737–18742. doi: 10.1073/pnas.1110552108

Zhang, Z., Mao, L., Chen, H., Bu, F., Li, G., Sun, J., et al. (2015). Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell* 27, 1595–1604. doi: 10.1105/tpc.114.135848

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876. doi: 10.1038/s41588-019-0402-2

Zohren, J., Wang, N., Kardailsky, I., Borrell, J. S., Joecker, A., Nichols, R. A., et al. (2016). Unidirectional diploid–tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Mol. Ecol.* 25, 2413–2426. doi: 10.1111/mec.13644

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 August 2020
doi: 10.3389/fgene.2020.00888

Transcriptome Dynamics of the Inflorescence in Reciprocally Formed Allopolyploid *Tragopogon miscellus* (Asteraceae)

Shengchen Shan[1,2], J. Lucas Boatwright[3], Xiaoxian Liu[4,5], Andre S. Chanderbali[2], Chaonan Fu[6], Pamela S. Soltis[1,2,7,8]* and Douglas E. Soltis[1,2,4,7,8]**

[1] Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL, United States, [2] Florida Museum of Natural History, University of Florida, Gainesville, FL, United States, [3] Advanced Plant Technology Program, Clemson University, Clemson, SC, United States, [4] Department of Biology, University of Florida, Gainesville, FL, United States, [5] Environmental Genomics and Systems Biology (EGSB), Biosciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [6] Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [7] Biodiversity Institute, University of Florida, Gainesville, FL, United States, [8] Genetics Institute, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:
Yves Van de Peer, Ghent University, Belgium

Reviewed by:
Raju Datla, Global Institute for Food Security (GIFS), Canada
Guanjing Hu, Iowa State University, United States

***Correspondence:**
Shengchen Shan
shan158538@ufl.edu
Pamela S. Soltis
psoltis@flmnh.ufl.edu
Douglas E. Soltis
dsoltis@ufl.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *29 May 2020*
Accepted: *20 July 2020*
Published: *06 August 2020*

Citation:
Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS and Soltis DE (2020) Transcriptome Dynamics of the Inflorescence in Reciprocally Formed Allopolyploid Tragopogon miscellus (Asteraceae). Front. Genet. 11:888. doi: 10.3389/fgene.2020.00888

Polyploidy is an important evolutionary mechanism and is prevalent among land plants. Most polyploid species examined have multiple origins, which provide genetic diversity and may enhance the success of polyploids. In some polyploids, recurrent origins can result from reciprocal crosses between the same diploid progenitors. Although great progress has been made in understanding the genetic consequences of polyploidy, the genetic implications of reciprocal polyploidization remain poorly understood, especially in natural polyploids. *Tragopogon* (Asteraceae) has become an evolutionary model system for studies of recent and recurrent polyploidy. Allotetraploid *T. miscellus* has formed reciprocally in nature with resultant distinctive floral and inflorescence morphologies (i.e., short- vs. long-liguled forms). In this study, we performed comparative inflorescence transcriptome analyses of reciprocally formed *T. miscellus* and its diploid parents, *T. dubius* and *T. pratensis*. In both forms of *T. miscellus*, homeolog expression of ∼70% of the loci showed vertical transmission of the parental expression patterns (i.e., parental legacy), and ∼20% of the loci showed biased homeolog expression, which was unbalanced toward *T. pratensis*. However, 17.9% of orthologous pairs showed different homeolog expression patterns between the two forms of *T. miscellus*. No clear effect of cytonuclear interaction on biased expression of the maternal homeolog was found. In terms of the total expression level of the homeologs studied, 22.6% and 16.2% of the loci displayed non-additive expression in short- and long-liguled *T. miscellus*, respectively. Unbalanced expression level dominance toward *T. pratensis* was observed in both forms of *T. miscellus*. Significantly, genes annotated as being involved in pectin catabolic processes were highly expressed in long-liguled *T. miscellus* relative to the short-liguled form, and the majority of these differentially expressed genes were transgressively down-regulated in short-liguled *T. miscellus*. Given the known role of these genes in cell expansion,

they may play a role in the differing floral and inflorescence morphologies of the two forms. In summary, the overall inflorescence transcriptome profiles are highly similar between reciprocal origins of *T. miscellus*. However, the dynamic homeolog-specific expression and non-additive expression patterns observed in *T. miscellus* emphasize the importance of reciprocal origins in promoting the genetic diversity of polyploids.

Keywords: homeolog, inflorescence, non-additive expression, polyploidy, reciprocal formation, *Tragopogon*, transcriptome

INTRODUCTION

Polyploidy, also known as whole-genome duplication (WGD), is a major evolutionary force in all eukaryotes (Otto, 2007; Soltis et al., 2014, 2016; Wendel, 2015; Van de Peer et al., 2017). WGD events are particularly prevalent and important in land plants, especially ferns and flowering plants (Leebens-Mack et al., 2019). Many ancient WGD events occur near the origin of several large angiosperm clades and are accompanied by the evolution of novel traits (Soltis et al., 2009; Soltis and Soltis, 2016), adaptation to dramatic environmental changes (Wu et al., 2019), and rapid niche differentiation (Baniaga et al., 2020).

Two types of polyploids are generally recognized: allopolyploids are formed by hybridization and chromosome doubling between two species, whereas autopolyploids are derived from genome duplication within a species (Grant, 1981). In both allopolyploids and autopolyploids, recent studies have illustrated the dynamic nature of polyploid genome evolution at multiple levels, including: chromosomal variation, gene loss, homeolog (duplicated gene copies following allopolyploidy) expression bias (see **Box 1** for all terminologies used in this article), non-additive gene expression (including expression level dominance and transgressive expression; **Box 1**), transposable element activation, and epigenetic changes (reviewed in Osborn et al., 2003; Doyle et al., 2008; Jackson and Chen, 2010; Mayfield et al., 2011; De Smet and Van de Peer, 2012; Madlung and Wendel, 2013; Buggs et al., 2014; Soltis et al., 2014; Yoo et al., 2014; Spoelhof et al., 2017; Wendel et al., 2018; Doyle and Coate, 2019). In addition, homeolog expression bias and expression level dominance can be either balanced or unbalanced (**Box 1**; Grover et al., 2012).

Significantly, most polyploids that have been investigated genetically at the population level have formed repeatedly (Soltis and Soltis, 1999; Soltis et al., 2004). That is, the same polyploid species has formed multiple times from genetically different diploid individuals. Because of the genetic diversity of the parents, independent assortment and recombination within polyploids, and subsequent gene flow among polyploid populations, multiple origins may have an important influence on the genetic diversity of polyploids.

As a special case of multiple origins, reciprocal origins of allopolyploids having similar nuclear genomes but distinct cytoplasmic genomes have been documented in various species, including *Aegilops* spp. (Meimberg et al., 2009), *Androsace brigantiaca* (Dixon et al., 2009), *Asplenium* spp. (Perrie et al., 2010; Sessa et al., 2018), *Brassica napus* (Song and Osborn, 1992), *Platanthera huronensis* (Wallace, 2003), *Polypodium hesperium* (Sigel et al., 2014), *Senecio* spp. (Kadereit et al., 2006), and *Tragopogon miscellus* (Ownbey and McCollum, 1953; Soltis and Soltis, 1989). Many reciprocally formed polyploids are divergent in morphology and/or geographic distribution. However, reciprocal polyploidization remains poorly studied, and its genetic impact in natural polyploid populations is still largely unknown.

Tragopogon (Asteraceae) is an outstanding natural system for studies of recent and recurrent allopolyploidy. Three diploid *Tragopogon* species ($2n = 12$), *T. dubius*, *T. pratensis*, and *T. porrifolius*, were introduced from Europe to the Pacific Northwest of North America in the early 1900s. Ownbey (1950) identified two new allotetraploids ($2n = 24$) native to the Palouse region of eastern Washington and adjacent Idaho, United States: *T. miscellus* and *T. mirus*. Both allotetraploids are only 90–100 years old (45–50 generations in these biennial plants) (Soltis et al., 2004). The parents of *T. miscellus* are *T. dubius* and *T. pratensis*, and those of *T. mirus* are *T. dubius* and *T. porrifolius* (**Figure 1**). Allotetraploids *T. miscellus* and *T. mirus* formed at least 21 and 11 times, respectively (Soltis et al., 2004; Symonds et al., 2010). Intriguingly, *T. miscellus* formed reciprocally with resultant distinct floral and inflorescence morphologies: those allotetraploids with *T. dubius* as the maternal parent have long ligules and open inflorescences, and those with *T. pratensis* as the maternal parent have short ligules and closed inflorescences (Ownbey, 1950; Ownbey and McCollum, 1953; Soltis and Soltis, 1989; **Figure 1**). Initial studies of the consequences of recurrent (including reciprocal) polyploidization in *T. miscellus* revealed similar patterns of stochastic homeolog loss and silencing among populations (including both short- and long-liguled forms) and a slight preferential loss of *T. dubius* homeologs (Tate et al., 2006, 2009; Buggs et al., 2009, 2011, 2012).

Allopolyploidy perturbs cytonuclear interactions, which may lead to biased homeolog expression and/or gene conversion toward the maternal parent in organelle-targeted nuclear genes (Gong et al., 2012, 2014; Sehrish et al., 2015; Sharbrough et al., 2017). Therefore, we hypothesize that organelle-targeted nuclear genes will display biased homeolog expression toward *T. dubius* and *T. pratensis* in long-liguled and short-liguled *T. miscellus*, respectively. In addition, unbalanced expression level dominance toward the maternal parent has been reported in many polyploids (reviewed in Yoo et al., 2014). If this hypothesis holds, the direction of unbalanced expression level dominance should be different between reciprocal origins of *T. miscellus*. Lastly, we attempt to address which gene(s) may be responsible for the

BOX 1 | Terminology of the gene expression profiles in allopolyploids used in this article.

Type of analysis	Terminology	Definition
Homeolog-specific expression analysis	Homeolog expression bias	Unequal expression levels of the two parental homeologs of a gene in the allopolyploid.
	Balanced homeolog expression bias	When multiple genes are examined, equivalent number of genes display homeolog expression bias toward each parent.
	Unbalanced homeolog expression bias	When multiple genes are examined, more genes show homeolog expression bias toward one parental genome than the other.
	Parental legacy	When comparing the allopolyploid and its diploid progenitors, the expression patterns of the diploid parents are vertically transmitted to the allopolyploid (**Figure 2A**, categories 1–3).
	Absence of homeolog expression bias	The diploid parents are differentially expressed at a given locus, but homeolog expression of that gene is unbiased in the allopolyploid (**Figure 2A**, categories 4 and 5).
	Novel homeolog expression	Homeolog expression of a gene in the allopolyploid is biased toward the diploid parent that displays an equal or lower expression level than the other diploid parent at that locus (**Figure 2A**, categories 6–9).
Total gene expression analysis (combining the expression levels of both homeologs of a gene)	Expression level dominance	One type of non-additive gene expression in which the expression level of a gene in the allopolyploid is equivalent to only one of the diploid progenitors.
	Balanced expression level dominance	When multiple genes are examined, equivalent number of genes display expression level dominance toward each parent.
	Unbalanced expression level dominance	When multiple genes are examined, more genes show expression level dominance toward one parental genome than the other.
	Transgressive expression	A second type of non-additive gene expression in which the gene expression level in the allopolyploid is higher (transgressive up-regulation) or lower (transgressive down-regulation) than in both diploid parents.

different inflorescence morphology between reciprocally formed *T. miscellus*. For example, the roles of *CYC2* clade genes in determining flower type and controlling ligule growth have been demonstrated in many species of Asteraceae (reviewed in Elomaa et al., 2018). We can ask, do the *Tragopogon CYC2* orthologs express differentially between short- and long-liguled *T. miscellus*?

To dissect the gene expression changes following reciprocal polyploidization, inflorescence transcriptome dynamics of short- and long-liguled *T. miscellus* and their diploid parents were compared across ∼12,000 loci. We examined homeolog-specific expression in both ligule forms of *T. miscellus* by employing a robust Bayesian Poisson-Gamma model following methods in Boatwright et al. (2018). In addition, non-additive gene expression was analyzed in reciprocally formed *T. miscellus*, and differentially expressed loci were identified between short- and long-liguled *T. miscellus*. Previous studies of recurrent origins of *T. miscellus* employed only from 10 to 144 loci and mostly focused on leaf tissues (Tate et al., 2006, 2009; Buggs et al., 2009, 2011, 2012). By examining a much larger gene sample in the inflorescence, we can gain a more comprehensive understanding of the impact of reciprocal polyploidization on transcriptome dynamics in these naturally formed young polyploids.

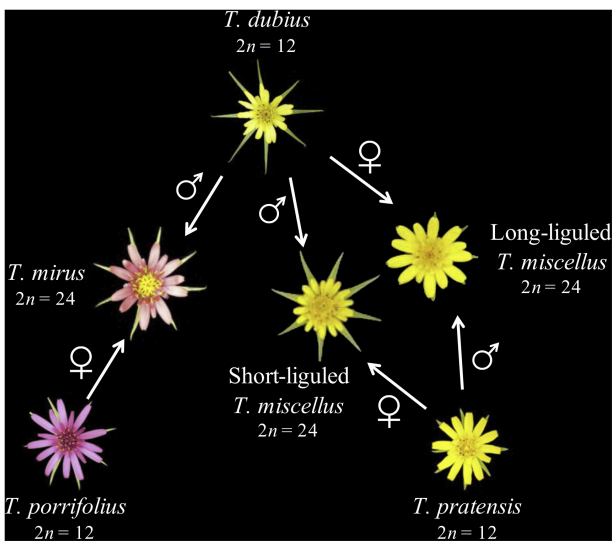


FIGURE 1 | Parentage of two recently formed allotetraploid species of *Tragopogon*. Polyploid *T. miscellus* is derived from *T. dubius* and *T. pratensis* and has formed reciprocally: when *T. pratensis* is the maternal parent, *T. miscellus* is short-liguled; the long-liguled form has *T. dubius* as the maternal progenitor. The maternal and paternal parents of allopolyploid *T. mirus* are *T. porrifolius* and *T. dubius*, respectively.

MATERIALS AND METHODS

Sample Collection and RNA Sequencing

Tragopogon plants were grown from field-collected seed in the Department of Biology greenhouse at the University of Florida, Gainesville, FL, United States. The following natural populations were analyzed in the current study (**Table 1**): *T. dubius* (Soltis and Soltis collection number 2613; Pullman, WA, United States), *T. dubius* (2886; Moscow, ID, United States), *T. pratensis* (2608; Moscow, ID, United States), short-liguled *T. miscellus* (2604; Moscow, ID, United States), and long-liguled *T. miscellus* (2605; Pullman, WA, United States). Because *T. pratensis* is extinct in Pullman, the paternal parent of long-liguled *T. miscellus* was represented by *T. pratensis* (2608) from Moscow in this study.

Herbarium vouchers for all of these samples were deposited in the Florida Museum of Natural History Herbarium (FLAS).

Fully opened inflorescences were collected from three individuals (one flower head per individual) of each sampled population and frozen in liquid nitrogen. RNA was extracted using a modified CTAB method (Jordon-Thaden et al., 2015). RNA-seq libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, United States). Libraries from *T. miscellus* and the diploid parents were sequenced using Illumina NextSeq and HiSeq, respectively, to generate paired-end 150-bp reads. All sequencing was performed at the Interdisciplinary Center for Biotechnology Research (ICBR), University of Florida, Gainesville, FL, United States.

Read Trimming

Approximately 135, 121, and 124 million raw paired-end reads were obtained from *T. dubius* (2613; Pullman), *T. dubius* (2886; Moscow), and *T. pratensis* (2608; Moscow), respectively (**Supplementary Table S1**). For *T. miscellus*, approximately 79 and 82 million raw paired-end reads were obtained from the short- and long-liguled form, respectively (**Supplementary Table S1**). Sequencing adaptors were removed using *Cutadapt* (version 2.1) (Martin, 2011). *Trimmomatic* (version 0.36) was used to remove low-quality bases (Bolger et al., 2014). Then, *sortmerna* (version 2.1) was used to remove rRNA sequences (Kopylova et al., 2012); 18S (accession number KT179662.1) and 26S (AF036493.1) ribosomal RNA genes from *T. dubius* were used as references.

Transcript Assembly and Redundancy Removal

Trinity (version r20180213-2.6.5) was used to assemble high-quality, rRNA-free reads from the diploid parental species (Haas et al., 2013). Normalized reads (maximum and minimum coverages are 50 and 2, respectively) were used for *de novo* transcriptome assembly with default parameters. The genotype concordance between the two *T. dubius* populations (Pullman and Moscow) was calculated by identifying sequence variation (SNP and indel) using *Picard* GenotypeConcordance[1] (**Supplementary Table S2**). Because of the high genotype concordance value, reads from the two populations of *T. dubius* were combined for the *T. dubius* transcript assembly (**Supplementary Table S2**). The quality of the assemblies was assessed by performing *BUSCO* analysis (Simão et al., 2015) and calculating ExN50 statistics (N50 value limited to the top highly expressed transcripts which account for x% of the total expression). In addition, by mapping reads back to assembled transcripts, the read composition of the assembly was assessed. To reduce assembly redundancy, *Lace* (Davidson et al., 2017) was used to generate SuperTranscripts, which comprise unique and common sequence regions from all isoforms derived from a single gene[2].

[1]http://broadinstitute.github.io/picard

TABLE 1 | Populations of *Tragopogon* analyzed (vouchers are deposited at FLAS).

Species	Soltis and Soltis collection number	Location
T. dubius	2613	Pullman, WA
	2886	Moscow, ID
T. pratensis	2608	Moscow, ID
T. miscellus (short-liguled)	2604	Moscow, ID
T. miscellus (long-liguled)	2605	Pullman, WA

Ortholog Calling Between Diploid Parents

Reciprocal best-hit orthologs were identified between SuperTranscripts of *T. dubius* and *T. pratensis* following Boatwright et al. (2018). In addition, *OrthoFinder* (version 2.3.3) was used to identify single-copy orthogroups between the two diploid species using peptide sequences predicted by *TransDecoder* (version 5.5.0) (Haas et al., 2013; Emms and Kelly, 2015). By comparing results from the above two approaches, shared orthologous pairs with high similarity (*E*-value $\leq$ 1e-10, identity $\geq$ 80%, alignment length $\geq$ 200 bp; BLAST results using SuperTranscripts from *T. pratensis* as query and those from *T. dubius* as database) were isolated and used to examine gene expression patterns in downstream analyses.

Homeolog-Specific Expression Analysis

The Poisson-Gamma model (León-Novelo et al., 2014) was used to assess homeolog-specific expression in the polyploids following Boatwright et al. (2018). Boatwright et al. (2018) emphasized the importance of assessing read mapping bias prior to performing homeolog-specific expression analysis. Briefly, diploid read alignments identified orthologous pairs showing mapping bias, i.e., the diploid reads from one parent were predominantly mapped to the reference from the other parent. Homeolog-specific expression was then examined in the remaining orthologous pairs that did not exhibit biased mapping. In the polyploids, each orthologous pair was classified as one of the following three possible categories: unbiased homeolog expression, biased homeolog expression toward *T. dubius*, and biased homeolog expression toward *T. pratensis* (**Figure 2B**). In addition, we examined the effect of parental gene expression level on the relative homeolog expression in the allopolyploid, which included three types of expression patterns: parental legacy (**Box 1** and **Figure 2A**, categories 1–3), absence of homeolog expression bias (**Box 1** and **Figure 2A**, categories 4 and 5), and novel homeolog expression bias (**Box 1** and **Figure 2A**, categories 6–9; Yoo et al., 2013, 2014).

Differential Gene Expression Analysis

Reads mapped to both of the 'common orthologous regions' between orthologous pairs were evaluated for differential gene expression (per Boatwright et al., 2018). Differential gene expression was analyzed using *DESeq2* (version 1.24.0) with the negative binomial generalized linear model (Love et al., 2014).

[2]https://github.com/trinityrnaseq/trinityrnaseq/wiki/SuperTranscripts

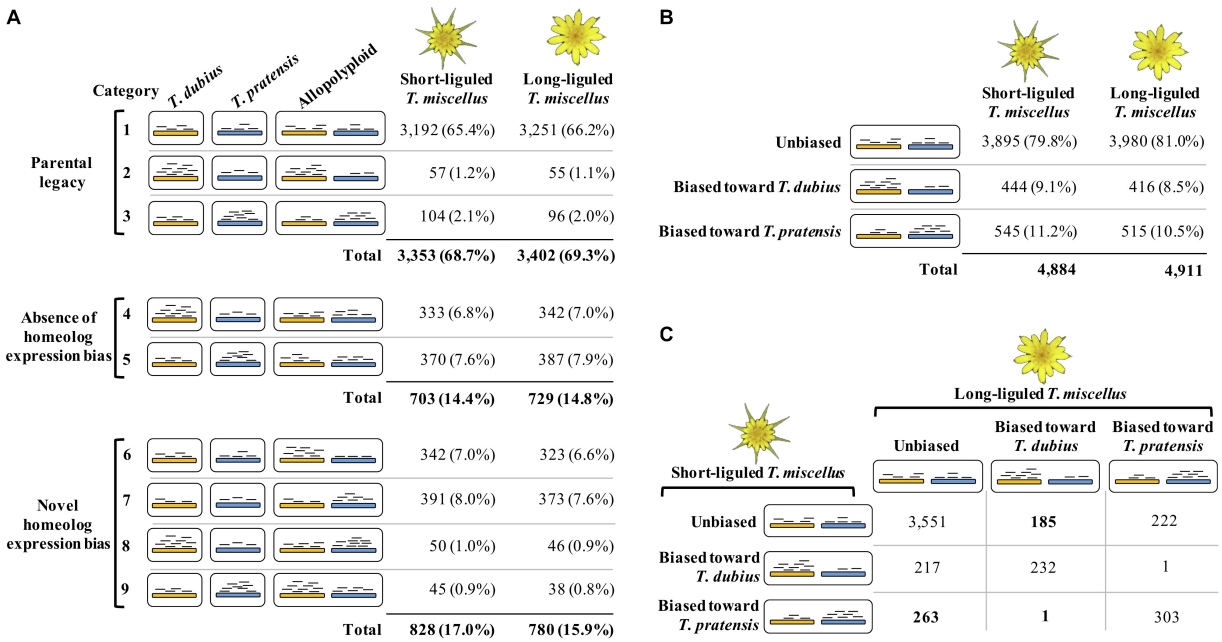


FIGURE 2 | Homeolog-specific expression analysis in short- and long-liguled *T. miscellus*. The orange and blue bars represent homeologs derived from *T. dubius* and *T. pratensis*, respectively. Black short lines indicate reads that are mapped to the reference — the abundance of reads indicates the expression level. **(A)** Effects of parental gene expression on homeolog-specific expression. Parental legacy (categories 1–3) indicates that the expression level of a gene in the diploid parents is vertically transmitted to the polyploid. For example, in category 2, a gene shows higher expression in *T. dubius* than in *T. pratensis*; in the polyploid, the gene has biased homeolog expression toward *T. dubius*. Homeolog expression bias can be absent in the polyploid (categories 4 and 5). For example, although a gene is differentially expressed between the two parents, in the allopolyploid, the expression levels of the two homeologs are equivalent. Novel homeolog expression bias (categories 6–9) is biased homeolog expression toward one parent that does not show higher expression than the other parent. For example, in category 6, a gene is not differentially expressed between the two diploid parents, but biased homeolog expression toward *T. dubius* is observed in the polyploid. **(B)** The number and proportion of loci showing unbiased and biased homeolog expression in short- and long-liguled *T. miscellus*. **(C)** Comparison of the homeolog-specific expression profiles between short- and long-liguled *T. miscellus*. The number of loci showing "lineage-specific biased homeolog expression toward the maternal parent" is displayed in bold.

Read count normalization in *DESeq2* takes both sequencing depth and RNA composition into consideration. For quality control, at the sample level, principal component analysis (PCA) and hierarchical clustering were performed, and any sample outliers were identified; at the gene level, those loci with zero total read counts, low mean-normalized counts, and extreme count outliers were removed from further analysis. If the adjusted *P*-value was below 0.05 [i.e., false discovery rate (FDR) is less than 5% (Benjamini and Hochberg, 1995)], the locus (gene) was considered to be differentially expressed. Gene expression levels were compared among *T. dubius*, *T. pratensis*, and short- and long-liguled *T. miscellus*. In addition, differential gene expression was assessed between the polyploid and the mid-parent value (MPV) of its diploid parents.

Analyses of Expression Level Dominance and Transgressive Expression

Differentially expressed loci between the polyploid and its diploid parents were parsed into various expression patterns (including additive expression, expression level dominance, and transgressive expression) according to Rapp et al. (2009) (**Figure 4A**). Briefly, if a gene was differentially expressed between the two diploids, additive expression indicates that the expression level in the polyploid is higher than one diploid parent but lower than the other one; a locus was classified as demonstrating expression level dominance when the expression level in the polyploid resembles that of one of the two parents; if the gene expression level in the polyploid was higher or lower than in both diploid parents, the locus was considered to show transgressive gene expression (**Box 1**). Following Yoo et al. (2013, 2014), in our study, non-additive expression includes both expression level dominance and transgressive expression (**Figure 4A**).

Trinotate Annotation

Trinotate (version 3.0.1) was used to annotate the SuperTranscripts and predicted protein sequences (Bryant et al., 2017). BLAST similarities were captured using BLASTx and BLASTp from the UniProt protein database (release 2019_06) (UniProt Consortium, 2015); based on the Pfam database (release 32.0) (Finn et al., 2016), *HMMER* (version 3.2.1) was used to identify protein domains (Zhang and Wood, 2003); signal peptides were predicted by *SignalP* (version 5.0b) (Petersen et al., 2011); transmembrane regions and rRNA transcripts were identified by running *tmHMM* (version 2.0c) (Krogh et al., 2001) and *RNAMMER* (version 1.2) (Lagesen et al., 2007), respectively. All annotation results were loaded into

a *Trinotate* SQLite Database (*E*-value threshold: 1e-5). Gene ontology (GO) assignments were obtained from UniProt and Pfam databases. Following Boatwright et al. (2018), the *GOseq* pipeline included in *Trinity* (version r20180213-2.6.5) was used for GO enrichment analysis by using GO terms derived from annotation of *T. dubius* assemblies (FDR < 0.05) (Young et al., 2010). The background gene set in the GO enrichment analysis included those loci used in assessing differential gene expression or homeolog-specific expression.

Data Availability

The raw sequence reads have been uploaded to the BioProject database from NCBI (BioProject ID: PRJNA633300). All scripts used in data analysis, the assemblies, and annotations are available at https://github.com/GatorShan/Tragopogon-Inflorescence-RNA-seq-Analysis.

RESULTS

Trinity Assembly and Redundancy Removal

After quality control, approximately 93 and 81% of reads remained in the diploid and polyploid samples, respectively (**Supplementary Table S1**). Using *de novo* assembly, in *T. dubius*, 126,278 *Trinity* genes and 302,750 *Trinity* transcripts were assembled with an N50 of 1,583 bp (based on all transcripts); GC percentage was 38.7%. In *T. pratensis*, the numbers of *Trinity* genes and *Trinity* transcripts were 99,228 and 239,956, respectively; GC percentage was 39.0%, and N50 was 1,482 bp (based on all transcripts).

BUSCO analysis was performed to assess completeness of the transcript assembly. Using 1,375 conserved single-copy orthologs from *Lactuca sativa* (Asteraceae) (a close relative of *Tragopogon*, in the same tribe; database: embryophyta_odb10) as the reference, 1,311 (95.3%) and 1,272 (92.5%) complete putative single-copy orthologs were identified in the *T. dubius* and *T. pratensis de novo* assemblies, respectively. In addition, when aligning RNA-seq reads back to *Trinity* assemblies, the overall alignment rate was 96.8% in both *T. dubius* and *T. pratensis*. Lastly, considered as a more appropriate criterion for transcriptome assembly evaluation than the N50 value, E90N50 statistics were computed by including the most highly expressed transcripts representing 90% of the total expression (Geniza and Jaiswal, 2017). E90N50 values were 1,802 and 1,663 bp for *T. dubius* and *T. pratensis de novo* assemblies, respectively. In summary, *de novo* assembly for the two diploid parents provided high-quality references for downstream gene expression analysis.

Lace was then used to remove isoform redundancy from the *Trinity de novo* assemblies: isoforms derived from a single gene were concatenated to create a SuperTranscript. For SuperTranscripts from *T. dubius*, the N50 value was 1,904 bp and the mean contig length was 921.8 bp. In *T. pratensis*, the N50 value and the average length of the SuperTranscripts were 1,954 and 981.7 bp, respectively. SuperTranscripts largely reduced redundancy of *Trinity* assemblies and were used to identify putative orthologs between the two diploid parents in the step noted below.

Ortholog Identification in *T. dubius* and *T. pratensis*

Using SuperTranscripts reconstructed from the previous step, 42,595 reciprocal best-hit orthologs were found between the two diploid parents following Boatwright et al. (2018). In addition, *OrthoFinder* identified 18,341 single-copy orthogroups (one protein sequence per species) shared between *T. dubius* and *T. pratensis*. Of the 12,900 shared loci between reciprocal best-hit and *OrthoFinder* results, 11,863 orthologous pairs with high confidence (*E*-value $\leq$ 1e-10, identify ≥ 0.8, and alignment length ≥ 200 bp) were isolated for downstream gene expression analysis.

Homeolog-Specific Expression in *T. miscellus*

Of 11,863 orthologous pairs identified in the two diploid parents, 5,400 loci showed unbiased mapping while aligning diploid reads to the SuperTranscripts. If reads from *T. dubius* mapped preferentially to the *T. pratensis* SuperTranscript, the orthologous pair was considered as showing mapping bias and was removed from the homeolog-specific expression analysis. Of the 5,400 unbiased orthologous pairs, 4,884 and 4,911 loci were examined in short- and long-liguled *T. miscellus*, respectively, along with their diploid parental populations. Three possible expression patterns are expected to be observed in polyploids: parental legacy, absence of homeolog expression bias, and novel homeolog expression bias (**Box 1**; **Figure 2A**).

The majority of orthologous pairs exhibited parental legacy in both forms of *T. miscellus* (68.7 and 69.3% in the short- and long-liguled forms, respectively) (**Figure 2A**, categories 1–3). For example, in short-liguled *T. miscellus*, the loci showing parental legacy included: (1) orthologous pairs that were not differentially expressed between *T. dubius* and *T. pratensis* and showed unbiased homeolog expression in the polyploid (65.4%); and (2) loci having biased homeolog expression toward the parent showing higher expression than that of the other parental species (3.3%) (**Figure 2A**). Following polyploidization, 14.4 and 14.8% of loci showed an absence of homeolog expression bias in short- and long-liguled *T. miscellus*, respectively (**Figure 2A**, categories 4 and 5). In addition, 17.0% of loci in short-liguled *T. miscellus* gained novel biased homeolog expression; in the long-liguled form, 15.9% of loci showed novel bias (**Figure 2A**, categories 6–9). Lastly, 81.9% of the loci examined (3,921 of 4,789) fell into the same category for both short- and long-liguled *T. miscellus* (**Supplementary Figure S1**).

Overall, 79.8 and 81.0% of orthologous pairs exhibited unbiased homeolog expression in short- and long-liguled *T. miscellus*, respectively (**Figure 2B**). In the short-liguled form, 444 (9.1%) and 545 (11.2%) loci showed biased homeolog expression toward *T. dubius* and *T. pratensis*, respectively (**Figure 2B**). In long-liguled *T. miscellus*, 416 (8.5%) loci displayed biased homeolog expression toward *T. dubius*, and the number of loci showing biased homeolog expression

toward *T. pratensis* was 515 (10.5%) (**Figure 2B**). In both short- and long-liguled *T. miscellus*, unbalanced homeolog expression bias toward *T. pratensis* was found to be significant (*P*-value equals 1.3e-3 and 1.2e-3 in short- and long-liguled *T. miscellus*, respectively; chi-square goodness-of-fit test).

We then asked whether the same orthologous pairs showed consistent homeolog-specific expression patterns between short- and long-liguled *T. miscellus*. Of the 4,975 loci we examined, 4,086 orthologous pairs (82.1%) showed the same homeolog-specific expression profiles between the short- and long-liguled forms (**Figure 2C**). For example, for 232 loci, both short- and long-liguled *T. miscellus* had biased homeolog expression toward *T. dubius* (**Figure 2C**). When the homeolog-specific expression profiles differed between reciprocally formed *T. miscellus*, in most cases, orthologous pairs showing biased homeolog expression in one form (either long- or short-liguled) showed unbiased homeolog expression in the other form. In long-liguled *T. miscellus*, there were 185 and 222 loci that showed biased homeolog expression toward *T. dubius* and *T. pratensis*, respectively, in contrast, all of these loci exhibited unbiased homeolog expression in short-liguled *T. miscellus* (**Figure 2C**). In very rare cases (only 2 of 4,975 loci), the direction of homeolog expression bias was altered between short- and long-liguled forms (e.g., bias toward *T. dubius* in long-liguled *T. miscellus*, but toward *T. pratensis* in the short-liguled form) (**Figure 2C**).

Differential Gene Expression Between *T. dubius* and *T. pratensis*

Principal component analysis and hierarchical clustering analysis indicated that all samples from the same diploid species clustered together (**Supplementary Figure S2**). In addition, the genotype concordance analysis between the Pullman and Moscow populations of *T. dubius* showed a high sequence similarity (**Supplementary Table S2**), and only one locus showed differential gene expression between the two populations of *T. dubius* (**Supplementary Figure S3**). Therefore, data from all *T. dubius* individuals (from Pullman and Moscow) were combined for differential gene expression analysis. After quality control, 10,746 orthologous pairs remained. Between the two diploid species, 2,017 (18.8%) loci showed differential gene expression, and 987 (9.2%) and 1,030 (9.6%) loci were highly expressed in *T. dubius* and *T. pratensis*, respectively (**Figure 3**). We did not find any enriched GO term in the differentially expressed loci between *T. dubius* and *T. pratensis*.

Differential Gene Expression Between Polyploids and Their Diploid Parents

Based on the hierarchical analysis, one individual of long-liguled *T. miscellus* (2605-9) did not cluster with the other two individuals from the same population (**Supplementary Figure S4**). In addition, the RNA integrity number (RIN) of individual 2605-9 was considerably lower than all other polyploid samples in this study (**Supplementary Table S3**). Because transcript quantification was likely affected by RNA integrity (Romero et al., 2014), this individual of long-liguled *T. miscellus* (2605-9) was removed in the differential gene expression analysis of this study. In terms of short-liguled *T. miscellus*, PCA and hierarchical analysis showed that all three replicates clustered together (**Supplementary Figure S4**).

The proportions of differentially expressed orthologous pairs between diploid parents and the two forms of *T. miscellus* are shown in **Figure 3**. Differential expression was found in 20.2% (2,128) of the loci examined between short-liguled *T. miscellus* and *T. dubius*; 17.5% (1,725) of orthologous pairs were differentially expressed between short-liguled *T. miscellus* and *T. pratensis* (**Figure 3**). In long-liguled *T. miscellus*, 15.0% (1,481) and 11.2% (1,103) of loci showed differential expression relative to *T. dubius* and *T. pratensis*, respectively (**Figure 3**).

Expression levels of orthologous pairs in polyploids were also compared with the mid-parent value (MPV) of the two diploid progenitors: 17.5% (1,768) and 11.2% (1,007) of loci were differentially expressed between the polyploid and the MPV in short- and long-liguled *T. miscellus*, respectively (**Figure 3**). In short-liguled *T. miscellus*, 8.0% (804) of loci were up-regulated relative to MPV, and 9.5% (964) of loci were down-regulated; in long-liguled *T. miscellus*, 5.5% (492) and 5.8% (515) of loci were up- and down-regulated compared to MPV, respectively (**Figure 3**).

Expression Level Dominance and Transgressive Expression in *T. miscellus*

By comparing the gene expression levels in the polyploid *T. miscellus* and its diploid parents, each locus was parsed into one of the expression patterns as shown in **Figure 4A**. In short- and long-liguled *T. miscellus*, 8,185 and 8,288 loci were examined, respectively. In the short-liguled form, the expression levels of 76.3% of the loci were not changed between the polyploid and its diploid parents; in the long-liguled form, 83.0% of the loci examined showed consistent expression between the polyploid and the diploid species (**Figure 4A**; expression pattern: no change). In addition, 1.1% of loci were additively expressed in short-liguled *T. miscellus*; in the long-liguled form, 0.7% of loci displayed additive expression (**Figure 4A**; expression pattern: additivity).

Non-additive expression, including expression level dominance and transgressive expression, was identified in 22.6 and 16.2% of the loci examined in short- and long-liguled *T. miscellus*, respectively (**Figure 4A**). Fisher's exact test indicated that the proportions of non-additive loci were significantly different between the short- and long-liguled forms (*P*-value $<$ 2.2e-16). In the short-liguled form, expression level dominance was unbalanced between the two subgenomes (*P*-value = 1.1e-2; chi-square goodness-of-fit test): 6.2 and 7.2% of loci showed expression level dominance toward *T. dubius* and *T. pratensis*, respectively (**Figure 4A**). In addition, in short-liguled *T. miscellus*, 5.9% of loci were transgressively down-regulated, and 3.3% of loci were transgressively up-regulated (**Figure 4A**). In the long-liguled form, expression level dominance toward *T. dubius* and *T. pratensis* was found in 4.8 and 6.3% of loci, respectively; a chi-square goodness-of-fit test showed that the expression level dominance was unbalanced

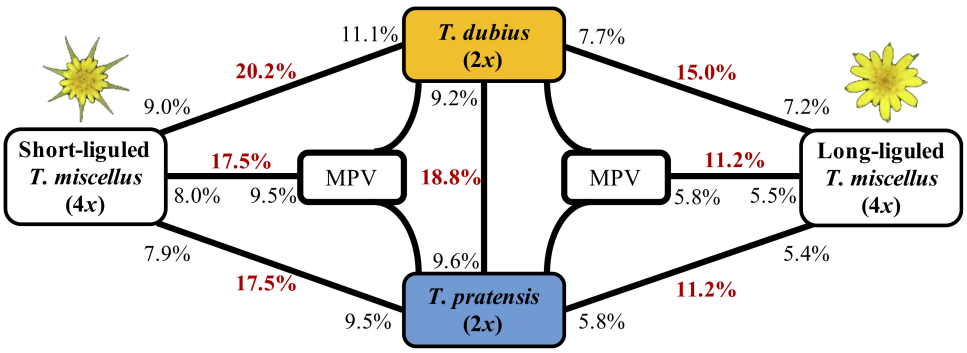


FIGURE 3 | Differentially expressed genes between *T. dubius*, *T. pratensis*, and reciprocally formed *T. miscellus*. The number in bold red indicates the fraction of genes that are differentially expressed in each contrast. The number in black indicates the proportion of up-regulated genes, and the placement of the number indicates the direction of up-regulation. For example, 20.2% of loci were differentially expressed between *T. dubius* and short-liguled *T. miscellus*: 11.1% of the loci examined showed higher expression in *T. dubius* than in short-liguled *T. miscellus*; 9.0% of loci were more highly expressed in short-liguled *T. miscellus* than in *T. dubius*. MPV = mid-parent value.

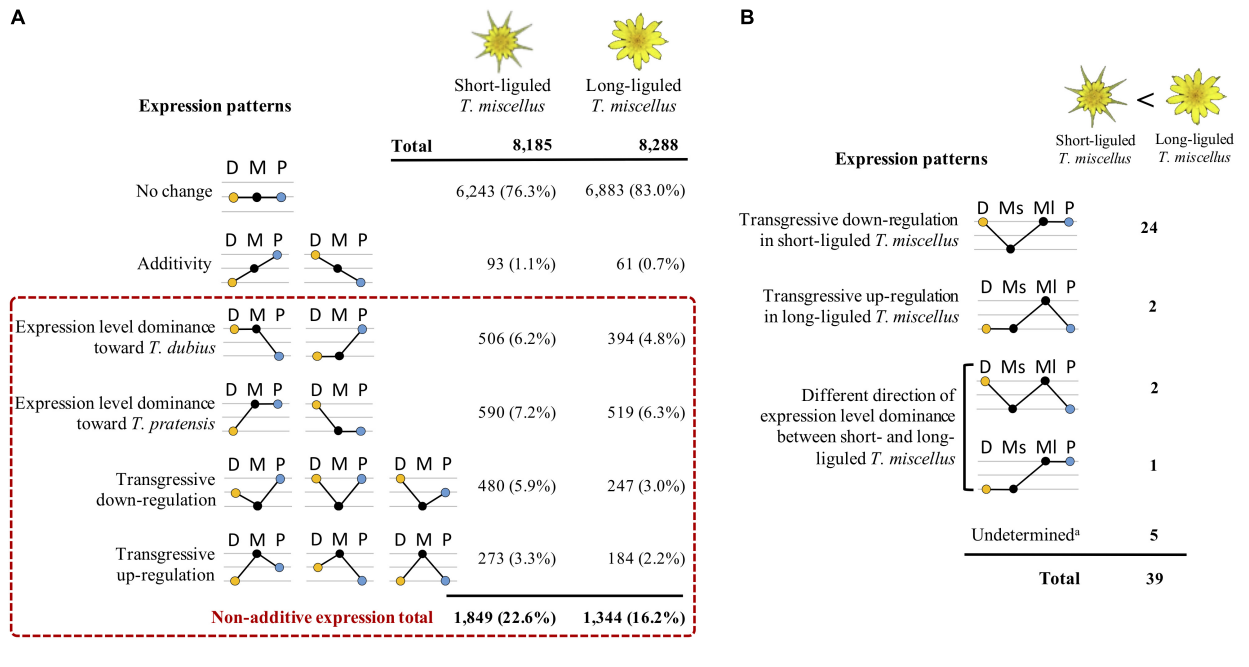


FIGURE 4 | Differential expression patterns in reciprocally formed allopolyploid *T. miscellus* relative to their diploid parents. **(A)** Comparison of the number of loci belonging to various expression patterns between short- and long-liguled *T. miscellus*. Non-additive expression includes expression level dominance and transgressive expression. D = *T. dubius*, M = *T. miscellus*, P = *T. pratensis*. **(B)** The effect of non-additive expression on the 39 loci showing higher expression in long-liguled *T. miscellus* than in the short-liguled form. Ms = short-liguled *T. miscellus*, Ml = long-liguled *T. miscellus*. a: this includes loci that display differential expression between the two forms of *T. miscellus* and yet show equivalent expression between the polyploid and the diploid progenitors when the expression level of each form of *T. miscellus* is individually compared with its diploid parents.

toward *T. pratensis* (*P*-value = 3.5e-5) (**Figure 4A**). Additionally, 3.0 and 2.2% of loci were transgressively down- and up-regulated in long-liguled *T. miscellus*, respectively (**Figure 4A**).

Gene ontology enrichment analysis was performed using transgressively expressed loci in the polyploids. In short-liguled *T. miscellus*, the transgressively down-regulated loci showed enriched GO terms (biological process) in pectin catabolic process and cell wall modification (**Supplementary Table S4**); for transgressively up-regulated loci, no enriched GO term was found. In long-liguled *T. miscellus*, we did not find any overrepresented GO term in the transgressively expressed loci.

Differential Gene Expression Between Reciprocally Formed *T. miscellus*

The transcriptomes of short- and long-liguled *T. miscellus* were compared, and 41 differentially expressed loci were identified. Of these loci, 39 orthologous pairs showed significantly higher expression in the long-liguled form than the short-liguled form;

GO analysis indicated overrepresentation of genes involved in two biological processes: pectin catabolic process and cell wall modification (**Supplementary Table S5**). For the two genes that were significantly highly expressed in short-liguled *T. miscellus*, no GO term was enriched.

In addition, we analyzed the relationship of non-additive expression to differential gene expression in the reciprocal origins of *T. miscellus*. Of the 39 loci showing higher expression in long-liguled *T. miscellus* than the short-liguled form, the majority of loci (24 of 39) were transgressively down-regulated in short-liguled *T. miscellus* (**Figure 4B**). That is, the expression levels of these 24 loci were not changed between long-liguled *T. miscellus* and its diploid parents, and because these loci showed transgressive down-regulation in the short-liguled form, the expression levels of the 24 loci were higher in long-liguled *T. miscellus* relative to the short-liguled form.

We also assessed the relationship of homeolog-specific expression to differential gene expression. Homeolog-specific expression was examined in 24 of the 39 loci showing higher expression in long-liguled *T. miscellus* than in the short-liguled form. The majority of these loci (15 of 24; 62.5%) displayed unbiased homeolog expression in both short- and long-liguled *T. miscellus* (**Supplementary Figure S5**). Considering the two loci showing higher expression in the short-liguled form relative to the long-liguled form, homeolog-specific expression was examined in one locus which also showed unbiased homeolog expression in both forms. Therefore, the up-regulated gene expression observed in one ligule form was mainly due to the increased expression levels of both homeologs.

DISCUSSION

Homeolog-Specific Expression in Polyploids

For each locus in the polyploid, homeolog-specific expression examines the relative contribution of homeologs derived from different diploid progenitors to the total gene expression. Biased homeolog expression might result from *cis*- and *trans*-regulatory changes (Wang et al., 2006a; Shi et al., 2012; Hu and Wendel, 2019) and/or DNA methylation dynamics: epigenetically silenced TEs can repress the expression of nearby homeologs, and variation in TE abundance and distribution between subgenomes may lead to homeolog expression bias (reviewed in Wendel et al., 2018). In the long run, homeologs with lower expression can be preferentially lost, which ultimately promotes biased fractionation (reviewed in Wendel et al., 2018). In addition, Yang et al. (2016) indicated that genes showing homeolog expression bias had more selection potential than genes displaying unbiased homeolog expression in polyploid *Brassica*, a phenomenon with potential agricultural applications for accelerating the breeding process.

Homeolog-specific expression has been studied in both natural and synthetic systems, including *Arabidopsis* (Wang et al., 2006a; Chang et al., 2010), *Brachypodium* (Takahagi et al., 2018), *Brassica* (Yang et al., 2016), *Coffea* (Combes et al., 2013), *Glycine* (Coate et al., 2014), *Gossypium* (Yoo et al., 2013; Rambani et al., 2014), *Mimulus* (Edger et al., 2017), *Polypodium* (Sigel et al., 2019), *Tragopogon* (Buggs et al., 2010; Boatwright et al., 2018), *Triticum* (Nomura et al., 2005), and *Zea* (Schnable et al., 2011).

Buggs et al. (2010) studied homeolog-specific expression in leaf tissue of short-liguled *T. miscellus* (2671; Oakesdale, WA, United States) and found that 69% of the loci examined showed equal expression of both homeologs. Also using leaf tissue, Boatwright et al. (2018) examined homeolog-specific expression of short-liguled *T. miscellus* from another population (2894-2; Garfield, WA, United States); the long-liguled form was not examined. Approximately 48% of loci showed unbiased homeolog expression, and 26.2 and 25.4% of orthologous pairs had biased homeolog expression toward *T. dubius* and *T. pratensis*, respectively (Boatwright et al., 2018). We found that, in the inflorescence of both short- and long-liguled *T. miscellus*, the majority of loci (∼80%) showed unbiased homeolog expression. In the short-liguled form, 9.1 and 11.2% of loci exhibited biased homeolog expression toward *T. dubius* and *T. pratensis*, respectively (**Figure 2B**).

As shown above, in short-liguled *T. miscellus*, the proportions of loci showing unbiased homeolog expression were divergent among studies. The difference may result from the method employed to assess homeolog-specific expression. In Buggs et al. (2010), the origin of a *T. miscellus* read was determined by using homeolog-specific single nucleotide polymorphism (SNP) markers between the diploid parents; instead, in Boatwright et al. (2018) and the current study, the analysis of homeolog-specific expression was based on a robust Bayesian Poisson-Gamma model for thousands of loci. In addition, different populations of short-liguled *T. miscellus* were analyzed in these studies, and divergent homeolog-specific expression profiles may be present at the population level. Lastly, Buggs et al. (2010) and Boatwright et al. (2018) examined homeolog-specific expression in leaf tissues, but the inflorescence transcriptome was analyzed in our study. The impact of tissue type on homeolog-specific expression is discussed below.

Homeolog-specific expression in allotetraploid *Gossypium hirsutum* has been examined using both leaf and petal transcriptomes (Yoo et al., 2013; Rambani et al., 2014): in cultivar "Maxxa," 73.2 and 79.4% of the genes examined showed unbiased homeolog expression in leaf and petal tissues, respectively; in a wild accession "TX2094," unbiased homeolog expression was found in 68.8% of genes in leaves, and 80.2% of the genes examined displayed equivalent expression of both homeologs in petals. In both leaf and petal tissues, biased homeolog expression was balanced (Yoo et al., 2013; Rambani et al., 2014). In allopolyploid *Brachypodium hybridum*, ∼60% of the genes showed unbiased homeolog expression in both leaf and root tissues (Takahagi et al., 2018). In summary, our results indicate that tissue type (leaf vs. inflorescence) may have an impact on relative homeolog expression.

Because *T. miscellus* is only 90–100 years old (45–50 generations) (Soltis et al., 2004), the extant diploid parental species are expected to have similar expression patterns compared to the exact ancestors of the polyploids. Therefore, the effects of parental gene expression profiles on relative homeolog expression in the polyploids can be rigorously

examined in *Tragopogon*, which is unique among various well-studied polyploid systems for this reason (Soltis and Soltis, 2009). In short- and long-liguled *T. miscellus*, 68.7 and 69.3% of orthologous pairs showed parental legacy (**Figure 2A**), indicating divergence from parental patterns at the remaining loci. In both forms of *T. miscellus*, approximately 15 and 16% of the loci examined displayed an absence of homeolog expression bias and novel homeolog expression bias in polyploids, respectively (**Figure 2A**). The percentage of loci showing parental legacy in *Tragopogon* is comparable to the results from other polyploid systems: ∼63 and ∼65% of analyzed genes showed parental legacy in natural allopolyploid *Gossypium* (Yoo et al., 2013) and allopolyploid *Brachypodium* (Takahagi et al., 2018), respectively. The fact that only approximately two-thirds of all loci examined in *T. miscellus* showed parental legacy after fewer than 50 generations of evolution of the polyploid indicates that deviations from parental patterns arise quickly after polyploid formation. This process occurs so rapidly that a young polyploid species such as *T. miscellus* has already undergone shifts in gene expression characteristic of much older polyploid species.

The Effect of Cytonuclear Interaction on Homeolog-Specific Expression

Cytonuclear incompatibility may occur following allopolyploidization: the nuclear genome is inherited from both parents, but the cytoplasmic genome is derived, typically, from the maternal parent (reviewed in Mogensen, 1996). Coordinated expression between cytoplasmic and nuclear genes is favored by selection following allopolyploidization. One intriguing potential compensatory mechanism is biased homeolog expression and/or gene conversion toward the maternal parent in organelle-targeted nuclear genes (Sharbrough et al., 2017). However, this hypothesis has only been tested with the *rbcS* gene (encodes the small subunit of Rubisco); asymmetric gene conversion biased toward the maternal copy has been reported (Gong et al., 2012, 2014).

Reciprocally formed *T. miscellus* provides an excellent system to examine the effect of cytonuclear interaction on homeolog-specific expression in allopolyploids. In this study, 17.9% of orthologous pairs (889 loci) showed different homeolog expression patterns between short- and long-liguled *T. miscellus* (**Figure 2C**). We isolated 449 loci with lineage-specific biased homeolog expression toward the maternal parent (**Figure 2C**, loci shown in bold). For example, this set of orthologous pairs included 185 loci showing biased homeolog expression toward *T. dubius* in long-liguled *T. miscellus* (of which *T. dubius* was the maternal parent), but unbiased homeolog expression in the short-liguled form. We hypothesized that if cytonuclear incompatibility resulted in biased homeolog expression toward the maternal parent, organelle-targeted nuclear genes should be included in these 449 loci and biological processes relevant to cytonuclear interactions (e.g., regulation of photosynthesis and oxidative phosphorylation) should be overrepresented in the GO analysis. However, no enriched GO term was found in these loci. In addition, genes with a mitochondrial/chloroplast transit peptide were not enriched in the 449 loci (**Supplementary Figure S6**). Consistent with our results, Sehrish et al. (2015) revealed that for the *rbcS* gene, a low proportion (16%) of naturally occurring *T. miscellus* individuals exhibited biased homeolog expression toward the maternal parent, and all of the synthetic *T. miscellus* individuals examined showed unbiased homeolog expression. Therefore, cytonuclear coordination may not be established immediately following polyploidization (Sehrish et al., 2015), and the number of loci showing biased homeolog expression toward the maternal parent is still low in the very young (90–100 years old) *T. miscellus*, which is, therefore, not captured by the GO analysis.

Non-additive Gene Expression in Polyploids

Non-additive gene expression, including expression level dominance and transgressive expression, is prevalent in diverse polyploid systems (reviewed in Yoo et al., 2014). The proportion of loci displaying non-additive expression varied significantly among different species, tissues, and environmental conditions. In allotetraploid *Arabidopsis suecica*, ∼5-6% of loci displayed non-additive expression (Wang et al., 2006b). In resynthesized allopolyploid wheat, non-additive expression was found in ∼19% of the loci examined (Akhunova et al., 2010). In *Gossypium hirsutum* cultivar "Maxxa," ∼29 and 61% of loci displayed non-additive expression in leaf and petal tissues, respectively (Yoo et al., 2013, 2014; Rambani et al., 2014). In allotetraploid *Coffea arabica*, ∼43 and 60% of loci showed non-additive expression under low and high temperatures, respectively (Bardil et al., 2011).

In our study, 22.6 and 16.2% of loci were non-additively expressed in short- and long-liguled *T. miscellus*, respectively (**Figure 4A**). These results mirrored those of the differential gene expression analysis between the polyploids and diploid progenitors: relative to the short-liguled form, fewer differentially expressed genes were found between the long-liguled form and either of the diploid parents (**Figure 3**).

In many polyploid species, unbalanced expression level dominance toward the maternal parent has been observed (reviewed in Yoo et al., 2014). For example, in leaf tissue of *Gossypium hirsutum* cultivar "Maxxa," 9.3 and 4.0% of loci showed expression level dominance toward the maternal parent (the A-genome) and paternal parent (the D-genome), respectively (Yoo et al., 2014). Similarly, in allopolyploid *Tragopogon mirus*, 8.5 and 7.8% of loci displayed expression level dominance toward *T. porrifolius* (the maternal parent) and *T. dubius* (the paternal parent), respectively (Yoo et al., unpublished).

The maternal influence on non-additive expression may result from: (1) the maternal inheritance of the cytoplasmic genome; and (2) cytonuclear incompatibilities (reviewed in Yoo et al., 2014; as described above in "The effect of cytonuclear interaction on homeolog-specific expression"). However, in the current study, irrespective of the cross direction, unbalanced expression level dominance toward *T. pratensis* was observed (**Figure 4A**). Consistent with these results, cytonuclear incompatibility did not result in biased homeolog expression toward the maternal

parent among organelle-targeted nuclear genes in *T. miscellus*. In allotetraploid *Polypodium hesperium*, unbalanced expression level dominance toward diploid *P. amorphum* was found in both reciprocal origins of the polyploid (Sigel et al., 2019). In addition, unbalanced expression level dominance toward the paternal parent was also found in polyploid *Coffea* (Bardil et al., 2011) and *Gossypium* (Rapp et al., 2009). More research is clearly needed on additional polyploid systems having reciprocal formations.

Interestingly, the transgressively down-regulated loci in short-liguled *T. miscellus* showed enriched GO terms in pectin catabolic process and cell wall modification (**Supplementary Table S4**); the exact two GO terms were overrepresented among loci showing higher expression in long-liguled *T. miscellus* than the short-liguled form (**Supplementary Table S5**) (the potential impact of these loci on morphology is discussed below). Consistent with these results, 24 of the 39 loci showing higher expression in long-liguled *T. miscellus* relative to the short-liguled form were transgressively down-regulated in short-liguled *T. miscellus* (**Figure 4B**). Therefore, the distinct non-additive expression patterns between reciprocal origins provide increased genetic diversity in the polyploids, which might lead to their success.

Distinctive Inflorescence Morphology Between Reciprocally Formed *T. miscellus*

Asteraceae are well known for their unique inflorescence (i.e., capitulum or flower head) — a key morphological innovation that is associated with their evolutionary success (Broholm et al., 2014). The flower head can be either homogamous (i.e., comprising one floral type) or heterogamous (i.e., comprising both ray and disk flowers). *Tragopogon* has flower heads composed of ligulate flowers only.

Previous studies have shown the important role of genes from the *CYC2* clade of the *CYCLOIDEA/TEOSINTE BRANCHED1* (*CYC/TB1*) gene family in defining floral identity and controlling ligule growth in other Asteraceae, including *Gerbera*, *Helianthus*, *Senecio*, and *Chrysanthemum* (reviewed in Elomaa et al., 2018). In developing flower primordia of *Gerbera* and *Helianthus*, *CYC2* clade genes exhibit higher expression in ray flower primordia than in disk flower primordia (Broholm et al., 2008; Tähtiharju et al., 2012). Disrupted expression of *CYC2* genes affects ligule length in ray and/or disk flowers in a species-specific manner (reviewed in Elomaa et al., 2018). In addition, at later stages of inflorescence development (including the fully opened inflorescence), *CYC2* clade genes are highly expressed in floral reproductive organs (stamen, stigma, style, and ovary) in *Gerbera* and *Helianthus* (Tähtiharju et al., 2012). However, how *CYC2* clade genes function in Asteraceae species with homogamous flower heads, such as *Tragopogon*, is still unknown.

We identified six SuperTranscripts (from *T. dubius*) that were homologous to the 13 *T. dubuis CYC/TB1* clade genes identified by Liu (2018) in a genome assembly of *T. dubius*. The expression profiles of these SuperTranscripts were examined, and none of them were differentially expressed between short- and long-liguled *T. miscellus* (**Supplementary Table S6**). Therefore, *CYC/TB1* clade genes may not be a factor in ligule length in *Tragopogon*. However, to better assess the role of *CYC2* genes in controlling ligule length in *Tragopogon*, future studies should also examine inflorescence transcriptomes from very early developmental stages — in *Gerbera* and *Helianthus*, it is in primordia that *CYC2* genes have shown differential expression between floral types with different ligule length (Elomaa et al., 2018). If the ligule length difference between reciprocally formed *T. miscellus* results from differential expression of *CYC2* orthologs in floral primordia, the differential expression of *CYC2* clade genes might not be captured by transcriptomes from inflorescences at anthesis (the material available and examined in this study). In addition, *CYC2* genes are expressed in reproductive organs at later inflorescence developmental stages in *Gerbera* and *Helianthus* (Tähtiharju et al., 2012). Therefore, to examine the function of *CYC2* genes in determining ligule growth, organ-specific transcriptomic studies of ligulate flowers would also be useful in future studies of *Tragopogon*.

In our study, 39 differentially expressed orthologous pairs were identified with higher expression in long-liguled *T. miscellus* compared to the short-liguled form. We are especially interested in the two significantly overrepresented GO terms of these loci: pectin catabolic process and cell wall modification (**Supplementary Table S5**). Pectin is a prominent component of the primary cell wall and the middle lamella. Numerous studies have indicated the important role of pectin modification in the regulation of cell wall extensibility, which has an impact on plant growth (reviewed in Palin and Geitmann, 2012; Wolf and Greiner, 2012). On the one hand, pectinesterase catalyzes the demethylesterification of pectin. In shoot apical meristems, pectin demethylesterification softens the cell wall and triggers primordium formation: overexpression of pectinesterase results in an increased number of primordia, and in turn, overexpression of pectinesterase inhibitor blocks primordium formation (Palin and Geitmann, 2012; Wolf and Greiner, 2012). In our study, pectinesterase activity is the most significantly overrepresented GO term in molecular function (FDR = 3.5e-3).

Another gene of interest encodes polygalacturonase. This enzyme cleaves demethylated pectin, which loosens the cell wall and enables cell expansion (Wolf and Greiner, 2012). Changes in polygalacturonase expression can perturb normal floral organ patterning in *Arabidopsis* (Xiao et al., 2014). A higher proportion of flowers had extra petals in both *PGX1* (*POLYGALACTURONASE INVOLVED IN EXPANSION1*) overexpression and mutant plants compared to that of the wild type counterparts (Xiao et al., 2014). In our study, pectin catabolic process is significantly enriched among differentially expressed loci showing higher expression in long-liguled *T. miscellus* than the short-liguled form (**Supplementary Table S5**).

Therefore, we speculate that the very few genes involved in pectin metabolism (especially genes controlling pectinesterase and polygalacturonase activity) may affect primordia formation and floral organ patterning in *Tragopogon*, and thereby contribute to the morphological differences in inflorescence structure between reciprocal formations of *T. miscellus* (short- vs.

long-liguled forms) — they should be a focus of more study. In plants, major morphological changes may be governed by just a few genes. For example, Hilu (1983) proposed the role of single-gene mutation in major morphological shifts in plants. Similarly, Gottlieb (1984) argued that very few genes might be responsible for structure, shape, and orientation changes. Doebley and Stec (1991) found that five genomic regions (one region included the *tb1* gene) could explain the dramatic inflorescence morphology difference between maize and teosinte. In a recent study of the carnivorous plant *Utricularia gibba*, the ectopic expression of a single gene (*UgPHV1*) impeded trap formation (Whitewoods et al., 2020).

Our hypothesis requires further study — an effect of pectin metabolism on inflorescence development has not been reported in Asteraceae. With an established CRISPR/Cas9 system in *Tragopogon* (Shan et al., 2018, 2020), the functions of related genes (such as genes encoding pectinesterase and polygalacturonase) could be rigorously examined.

In addition, there are two other important directions for future studies. First, the very few differentially expressed loci between the two ligule forms identified in our study may be targeted by a single transcription factor (TF). To test this hypothesis, the promoter sequences of the differentially expressed genes can be analyzed to identify any TF binding site(s). Second, the effect of the cytoplasm and cytonuclear interaction on ligule growth should be examined. Ownbey and McCollum (1953) crossed the diploid species of *Tragopogon* and assessed the inflorescence phenotype of the F_1 and F_2 generations. They found that ligule development was inhibited in the F_1 hybrids with the *T. pratensis* cytoplasm, and the subsequent F_2 generation displayed remarkable variation in the degree of ligule inhibition (Ownbey and McCollum, 1953). Therefore, future studies could compare the genotype and expression level of organellar genes (from plastids and mitochondria) between *T. dubius* and *T. pratensis*, and analyze the gene expression networks between organellar and nuclear genes in both diploids and their polyploid derivative (*T. miscellus*). All of these efforts will help provide a better understanding of the molecular mechanisms underlying ligule growth in *Tragopogon*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/, BioProject ID: PRJNA633300.

AUTHOR CONTRIBUTIONS

DS, PS, and SS designed the research. SS, XL, and CF collected the materials. SS, XL, and AC extracted RNA and constructed the libraries. SS and JB analyzed the data. SS drafted the manuscript. DS, PS, JB, XL, and AC participated in the writing of the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the US National Science Foundation (award number IOS-1923234 to DS and PS), the Florida Museum of Natural History, and a Botanical Society of America Graduate Student Research Award (to SS).

ACKNOWLEDGMENTS

We appreciate the help of Matthew A. Gitzendanner and Evgeny V. Mavrodiev on data analysis and laboratory work, respectively. We thank Mi-Jeong Yoo, Guanjing Hu, and Corrinne E. Grover for their helpful discussions. We also thank the reviewers for their helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00888/full#supplementary-material

REFERENCES

Akhunova, A. R., Matniyazov, R. T., Liang, H., and Akhunov, E. D. (2010). Homoeolog-specific transcriptional bias in allopolyploid wheat. *BMC Genomics* 11:505. doi: 10.1186/1471-2164-11-505

Baniaga, A. E., Marx, H. E., Arrigo, N., and Barker, M. S. (2020). Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecol. Lett.* 23, 68–78. doi: 10.1111/ele.13402

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

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Boatwright, J. L., McIntyre, L. M., Morse, A. M., Chen, S., Yoo, M. J., Koh, J., et al. (2018). A robust methodology for assessing differential homeolog contributions to the transcriptomes of allopolyploids. *Genetics* 210, 883–894. doi: 10.1534/genetics.118.301564

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Broholm, S. K., Tähtiharju S., Laitinen R. A. E., Albert V. A., Teeri T. H., and Elomaa P. (2008). A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9117–9122. doi: 10.1073/pnas.0801359105

Broholm, S. K., Teeri, T. H., and Elomaa, P. (2014). "Molecular control of inflorescence development in Asteraceae," in *Advances in Botanical Research: The Molecular Genetics of Floral Transition and Flower Development*, ed. F. Fornara (London, UK: Academic Press), 297–333.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063

Buggs, R. J. A., Chamala, S., Wu, W., Gao, L., May, G. D., Schnable, P. S., et al. (2010). Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol. Ecol.* 19, 132–146. doi: 10.1111/j.1365-294X.2009.04469.x

Buggs, R. J. A., Chamala, S., Wu, W., Tate, J. A., Schnable, P. S., Soltis, D. E., et al. (2012). Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol.* 22, 248–252. doi: 10.1016/j.cub.2011.12.027

Buggs, R. J. A., Doust, A. N., Tate, J. A., Koh, J., Soltis, K., Feltus, F. A., et al. (2009). Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* 103, 73–81. doi: 10.1038/hdy.2009.24

Buggs, R. J. A., Wendel, J. F., Doyle, J. J., Soltis, D. E., Soltis, P. S., and Coate, J. E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130354. doi: 10.1098/rstb.2013.0354

Buggs, R. J. A., Zhang, L., Miles, N., Tate, J. A., Gao, L., Wei, W., et al. (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* 21, 551–556. doi: 10.1016/j.cub.2011.02.016

Chang, P. L., Dilkes, B. P., McMahon, M., Comai, L., and Nuzhdin, S. V. (2010). Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biol.* 11:R125. doi: 10.1186/gb-2010-11-12-r125

Coate, J. E., Bar, H., and Doyle, J. J. (2014). Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *Plant Cell* 26, 136–150. doi: 10.1105/tpc.113.119966

Combes, M. C., Dereeper, A., Severac, D., Bertrand, B., and Lashermes, P. (2013). Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol.* 200, 251–260. doi: 10.1111/nph.12371

Davidson, N. M., Hawkins, A. D. K., and Oshlack, A. (2017). SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. *Genome Biol.* 18:148. doi: 10.1186/s13059-017-1284-1

De Smet, R., and Van de Peer, Y. (2012). Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* 15, 168–176. doi: 10.1016/j.pbi.2012.01.003

Dixon, C. J., Schönswetter, P., Suda, J., Wiedermann, M. M., and Schneeweiss, G. M. (2009). Reciprocal Pleistocene origin and postglacial range formation of an allopolyploid and its sympatric ancestors (*Androsace adfinis* group, Primulaceae). *Mol. Phylogenet. Evol.* 50, 74–83. doi: 10.1016/j.ympev.2008.10.009

Doebley, J., and Stec, A. (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129, 285–295.

Doyle, J. J., and Coate, J. E. (2019). Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* 180, 1–52. doi: 10.1086/700636

Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Elomaa, P., Zhao, Y., and Zhang, T. (2018). Flower heads in Asteraceae-recruitment of conserved developmental regulators to control the flower-like inflorescence architecture. *Hortic. Res.* 5:36. doi: 10.1038/s41438-018-0056-8

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Geniza, M., and Jaiswal, P. (2017). Tools for building *de novo* transcriptome assembly. *Curr. Plant Biol.* 11–12, 41–45. doi: 10.1016/j.cpb.2017.12.004

Gong, L., Olson, M., and Wendel, J. F. (2014). Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol. Biol. Evol.* 31, 2624–2636. doi: 10.1093/molbev/msu207

Gong, L., Salmon, A., Yoo, M. J., Grupp, K. K., Wang, Z., Paterson, A. H., et al. (2012). The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. *Mol. Biol. Evol.* 29, 3023–3036. doi: 10.1093/molbev/mss110

Gottlieb, L. D. (1984). Genetics and morphological evolution in plants. *Am. Nat.* 123, 681–709. doi: 10.1086/284231

Grant, V. (1981). *Plant Speciation*, 2nd Edn. New York, NY: Columbia University Press.

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hilu, K. W. (1983). "The role of single-gene mutations in the evolution of flowering plants," in *Evolutionary Biology*, eds M. K. Hecht, B. Wallace, and G. T. Prance (Boston, MA: Plenum Press), 97–128.

Hu, G., and Wendel, J. F. (2019). *Cis-trans* controls and regulatory novelty accompanying allopolyploidization. *New Phytol.* 221, 1691–1700. doi: 10.1111/nph.15515

Jackson, S., and Chen, Z. J. (2010). Genomic and expression plasticity of polyploidy. *Curr. Opin. Plant Biol.* 13, 153–159. doi: 10.1016/j.pbi.2009.11.004

Jordon-Thaden, I. E., Chanderbali, A. S., Gitzendanner, M. A., and Soltis, D. E. (2015). Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta. *Appl. Plant Sci.* 3:1400105. doi: 10.3732/apps.1400105

Kadereit, J. W., Uribe-Convers, S., Westberg, E., and Comes, H. P. (2006). Reciprocal hybridization at different times between *Senecio flavus* and *Senecio glaucus* gave rise to two polyploid species in north Africa and south-west Asia. *New Phytol.* 169, 431–441. doi: 10.1111/j.1469-8137.2005.01604.x

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

León-Novelo, L. G., McIntyre, L. M., Fear, J. M., and Graze, R. M. (2014). A flexible Bayesian method for detecting allelic imbalance in RNA-seq data. *BMC Genomics* 15:920. doi: 10.1186/1471-2164-15-920

Liu, X. (2018). *Alternative Splicing in Basal Angiosperms and Tragopogon (Asteraceae)*. Dissertation, University of Florida, Gainesville, FL.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Madlung, A., and Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet. Genome Res.* 140, 270–285. doi: 10.1159/000351430

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Mayfield, D., Chen, Z. J., and Pires, J. C. (2011). Epigenetic regulation of flowering time in polyploids. *Curr. Opin. Plant Biol.* 14, 174–178. doi: 10.1016/j.pbi.2011.03.008

Meimberg, H., Rice, K. J., Milan, N. F., Njoku, C. C., and McKay, J. K. (2009). Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *Am. J. Bot.* 96, 1262–1273. doi: 10.3732/ajb.0800345

Mogensen, H. L. (1996). The hows and whys of cytoplasmic inheritance in seed plants. *Am. J. Bot.* 83, 383–404. doi: 10.1002/j.1537-2197.1996.tb12718.x

Nomura, T., Ishihara, A., Yanagita, R. C., Endo, T. R., and Iwamura, H. (2005). Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16490–16495. doi: 10.1073/pnas.0505156102

Osborn, T. C., Pires, J. C., Birchler, J. A., Auger, D. L., Chen, Z. J., Lee, H. S., et al. (2003). Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19, 141–147. doi: 10.1016/s0168-9525(03)00015-5

Otto, S. P. (2007). The evolutionary consequences of polyploidy. *Cell* 131, 452–462. doi: 10.1016/j.cell.2007.10.022

Ownbey, M. (1950). Natural hybridization and amphiploidy in the genus *Tragopogon*. *Am. J. Bot.* 37, 487–499. doi: 10.2307/2438023

Ownbey, M., and McCollum, G. D. (1953). Cytoplasmic inheritance and reciprocal amphiploidy in *Tragopogon*. *Am. J. Bot.* 40, 788–796. doi: 10.2307/2438276

Palin, R., and Geitmann, A. (2012). The role of pectin in plant morphogenesis. *Biosystems* 109, 397–402. doi: 10.1016/j.biosystems.2012.04.006

Perrie, L. R., Shepherd, L. D., De Lange, P. J., and Brownsey, P. J. (2010). Parallel polyploid speciation: distinct sympatric gene-pools of recurrently derived allo-octoploid *Asplenium* ferns. *Mol. Ecol.* 19, 2916–2932. doi: 10.1111/j.1365-294X.2010.04705.x

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

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

Rapp, R. A., Udall, J. A., and Wendel, J. F. (2009). Genomic expression dominance in allopolyploids. *BMC Biol.* 7:18. doi: 10.1186/1741-7007-7-18

Romero, I. G., Pai, A. A., Tung, J., and Gilad, Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* 12:42. doi: 10.1186/1741-7007-12-42

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4069–4074. doi: 10.1073/pnas.1101368108

Sehrish, T., Symonds, V. V., Soltis, D. E., Soltis, P. S., and Tate, J. A. (2015). Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One* 10:e0144339. doi: 10.1371/journal.pone.0144339

Sessa, E. B., Vicent, M., and Chambers, S. M. (2018). Evolution and reciprocal origins in mediterranean ferns: the *Asplenium obovatum* and *A. adiantumnigrum* Complexes. *Ann. Missour. Bot. Gard.* 103, 175–187. doi: 10.3417/2018108

Shan, S., Mavrodiev, E. V., Li, R., Zhang, Z., Hauser, B. A., Soltis, P. S., et al. (2018). Application of CRISPR/Cas9 to *Tragopogon* (Asteraceae), an evolutionary model for the study of polyploidy. *Mol. Ecol. Resour.* 18, 1427–1443. doi: 10.1111/1755-0998.12935

Shan, S., Soltis, P. S., Soltis, D. E., and Yang, B. (2020). Considerations in adapting CRISPR/Cas9 in nongenetic model plant systems. *Appl. Plant Sci.* 8:e11314. doi: 10.1002/aps3.11314

Sharbrough, J., Conover, J. L., Tate, J. A., Wendel, J. F., and Sloan, D. B. (2017). Cytonuclear responses to genome doubling. *Am. J. Bot.* 104, 1277–1280. doi: 10.3732/ajb.1700293

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

Sigel, E. M., Der, J. P., Windham, M. D., and Pryer, K. M. (2019). Expression level dominance and homeolog expression bias in recurrent origins of the allopolyploid fern *Polypodium hesperium*. *Am. Fern J.* 109, 224–247. doi: 10.1640/0002-8444-109.3.224

Sigel, E. M., Windham, M. D., and Pryer, K. M. (2014). Evidence for reciprocal origins in *Polypodium hesperium* (Polypodiaceae): a fern model system for investigating how multiple origins shape allopolyploid genomes. *Am. J. Bot.* 101, 1476–1485. doi: 10.3732/ajb.1400190

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079

Soltis, D. E., and Soltis, P. S. (1989). Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. *Am. J. Bot.* 76, 1119–1124. doi: 10.2307/2444824

Soltis, D. E., and Soltis, P. S. (1999). Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* 14, 348–352. doi: 10.1016/s0169-5347(99)01638-9

Soltis, D. E., Soltis, P. S., Pires, J. C., Kovarik, A., Tate, J. A., and Mavrodiev, E. (2004). Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biol. J. Linn. Soc.* 82, 485–501. doi: 10.1111/j.1095-8312.2004.00335.x

Soltis, D. E., Visger, C. J., Marchant, D. B., and Soltis, P. S. (2016). Polyploidy: pitfalls and paths to a paradigm. *Am. J. Bot.* 103, 1146–1166. doi: 10.3732/ajb.1500501

Soltis, D. E., Visger, C. J., and Soltis, P. S. (2014). The polyploidy revolution then…and now: stebbins revisited. *Am. J. Bot.* 101, 1057–1078. doi: 10.3732/ajb.1400178

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Soltis, P. S., and Soltis, D. E. (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, K., and Osborn, T. C. (1992). Polyphyletic origins of *Brassica napus*: new evidence based on organelle and nuclear RFLP analyses. *Genome* 35, 992–1001. doi: 10.1139/g92-152

Spoelhof, J. P., Soltis, P. S., and Soltis, D. E. (2017). Pure polyploidy: closing the gaps in autopolyploid research. *J. Syst. Evol.* 55, 340–352. doi: 10.1111/jse.12253

Symonds, V. V., Soltis, P. S., and Soltis, D. E. (2010). Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow, and population structure. *Evolution* 64, 1984–2003. doi: 10.1111/j.1558-5646.2010.00978.x

Tähtiharju, S., Rijpkema, A. S., Vetterli, A., Albert, V. A., Teeri, T. H., and Elomaa, P. (2012). Evolution and diversification of the *CYC/TB1* gene family in Asteraceae – a comparative study in Gerbera (Mutisieae) and sunflower (Heliantheae). *Mol. Biol. Evol.* 29, 1155–1166. doi: 10.1093/molbev/msr283

Takahagi, K., Inoue, K., Shimizu, M., Uehara-Yamaguchi, Y., Onda, Y., and Mochida, K. (2018). Homoeolog-specific activation of genes for heat acclimation in the allopolyploid grass *Brachypodium hybridum*. *Gigascience* 7:giy020. doi: 10.1093/gigascience/giy020

Tate, J. A., Joshi, P., Soltis, K. A., Soltis, P. S., and Soltis, D. E. (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

Tate, J. A., Ni, Z., Scheen, A. C., Koh, J., Gilbert, C. A., Lefkowitz, D., et al. (2006). Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173, 1599–1611. doi: 10.1534/genetics.106.057646

UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

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

Wallace, L. E. (2003). Molecular evidence for allopolyploid speciation and recurrent origins in *Platanthera huronensis* (Orchidaceae). *Int. J. Plant Sci.* 164, 907–916. doi: 10.1086/378658

Wang, J., Tian, L., Lee, H. S., and Chen, Z. J. (2006a). Nonadditive regulation of *FRI* and *FLC* loci mediates flowering-time variation in *Arabidopsis* allopolyploids. *Genetics* 173, 965–974. doi: 10.1534/genetics.106.056580

Wang, J., Tian, L., Lee, H. S., Wei, N. E., Jiang, H., Watson, B., et al. (2006b). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172, 507–517. doi: 10.1534/genetics.105.047894

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Whitewoods, C. D., Gonçalves, B., Cheng, J., Cui, M., Kennaway, R., Lee, K., et al. (2020). Evolution of carnivorous traps from planar leaves through simple shifts in gene expression. *Science* 367, 91–96. doi: 10.1126/science.aay5433

Wolf, S., and Greiner, S. (2012). Growth control by cell wall pectins. *Protoplasma* 249, 169–175. doi: 10.1007/s00709-011-0371-5

Wu, S., Han, B., and Jiao, Y. (2019). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant.* 13, 59–71. doi: 10.1016/j.molp.2019.10.012

Xiao, C., Somerville, C., and Anderson, C. T. (2014). POLYGALACTURONASE INVOLVED IN EXPANSION1 functions in cell elongation and flower development in *Arabidopsis*. *Plant Cell* 26, 1018–1035. doi: 10.1105/tpc.114.123968

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* 48, 1225–1232. doi: 10.1038/ng.3657

Yoo, M. J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Yoo, M. J., Szadkowski, E., and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14

Zhang, Z., and Wood, W. I. (2003). A profile hidden Markov model for signal peptides generated by HMMER. *Bioinformatics* 19, 307–308. doi: 10.1093/bioinformatics/19.2.307

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 28 August 2020
doi: 10.3389/fgene.2020.01014

Homoeologous Exchanges, Segmental Allopolyploidy, and Polyploid Genome Evolution

Annaliese S. Mason[1] and Jonathan F. Wendel[2]*

[1] Plant Breeding Department, Justus Liebig University Giessen, Giessen, Germany, [2] Ecology, Evolution, and Organismal Biology Department, Iowa State University, Ames, IA, United States

Polyploidy is a major force in plant evolution and speciation. In newly formed allopolyploids, pairing between related chromosomes from different subgenomes (homoeologous chromosomes) during meiosis is common. The initial stages of allopolyploid formation are characterized by a spectrum of saltational genomic and regulatory alterations that are responsible for evolutionary novelty. Here we highlight the possible effects and roles of recombination between homoeologous chromosomes during the early stages of allopolyploid stabilization. Homoeologous exchanges (HEs) have been reported in young allopolyploids from across the angiosperms. Although all lineages undergo karyotype change via chromosome rearrangements over time, the early generations after allopolyploid formation are predicted to show an accelerated rate of genomic change. HEs can also cause changes in allele dosage, genome-wide methylation patterns, and downstream phenotypes, and can hence be responsible for speciation and genome stabilization events. Additionally, we propose that fixation of duplication – deletion events resulting from HEs could lead to the production of genomes which appear to be a mix of autopolyploid and allopolyploid segments, sometimes termed "segmental allopolyploids." We discuss the implications of these findings for our understanding of the relationship between genome instability in novel polyploids and genome evolution.

Keywords: polyploidy, homoeologous exchanges, chromosome behavior, synthetics, genome evolution

OPEN ACCESS

Edited by:
Jianping Wang,
University of Florida, United States

Reviewed by:
J. Chris Pires,
University of Missouri, United States
Robert VanBuren,
Michigan State University, United States
Pamela Soltis,
University of Florida, United States
Michael R. McKain,
The University of Alabama, United States

****Correspondence:***
Annaliese S. Mason
annaliese.mason@agrar.uni-giessen.de;
annaliese.mason@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *20 May 2020*
Accepted: *10 August 2020*
Published: *28 August 2020*

Citation:
Mason AS and Wendel JF (2020) Homoeologous Exchanges, Segmental Allopolyploidy, and Polyploid Genome Evolution. Front. Genet. 11:1014. doi: 10.3389/fgene.2020.01014

INTRODUCTION

Recent technological advances have vastly expanded access to genomic information, even for complex genomes (reviewed by Yuan et al., 2017). As an adjunct to *de novo* genome assembly for the creation of reference genomes, population genomic studies enable resequencing of multiple individuals within species to provide genetic data on a scale and at a resolution only dreamed of just a few years ago. In this new genomics era, it seems timely to revisit some of the fundamental concepts established in the early years of cytogenetics, particularly regarding insights into meiosis in polyploids and how this new understanding helps predict and explain several aspects of polyploid evolution and diversification.

In this review, we provide an overview of the cytogenetic processes associated with polyploidy, particularly the early stages of polyploid formation, and how these processes may induce genomic structural variation and give rise to novel phenotypes, thus providing an evolutionary

substrate for diversification. We revisit the idea that homoeologous recombination in polyploids may lead to rapid karyotypic and genomic restructuring in the first few generations after polyploid formation (Song et al., 1995), and in the process generate duplicated genomic regions and other findings that may seem difficult to explain based on species relationships and phylogenetic inferences. We further discuss how homoeologous exchanges can affect phenotypes to further impact the process of speciation via saltational changes. Lastly, we introduce the concept that homoeologous exchanges may be responsible for the observation of "segmental allopolyploidy," where auto- and allopolyploidy appear to both be present across the polyploid genome (Stebbins, 1950; Sybenga, 1996).

THE POLYPLOID SPECTRUM: FROM AUTO TO ALLO

Polyploidy, where three or more haploid chromosome sets are present within a single organism, is ubiquitous across the plant and animal kingdoms, with the minor exception of mammal and bird lineages (Van De Peer et al., 2017). Polyploids were first classified into "autopolyploids" and "allopolyploids" nearly a century ago by Kihara and Ono (1926), who proposed the distinction that autopolyploids derive from chromosome doubling of a single individual, and allopolyploids derive from hybridization. However, although chromosome doubling within reproductive tissue of a single individual may yield two identical chromosome complements, the frequency of this route to polyploidy remains unclear. As early as 1947, Stebbins cast doubt on the existence of natural autopolyploids formed via chromosome doubling (Stebbins, 1947). We now know that newly formed polyploids that arise via chromosome doubling are likely to suffer inbreeding depression (Abel and Becker, 2007), with major advantages conferred by heterozygosity in both auto- and allopolyploid species (for review see Bingham, 1980). Hence, it seems likely that most autopolyploidy events actually occur via sexual reproduction between two individuals within a species (for review see Soltis et al., 2014; Spoelhof et al., 2017b), or at the very least via meiotic events which allow for the generation of novel variation in progeny (for review see De Storme and Mason, 2014). In fact, the mechanism of "hybridization followed by chromosome doubling" was re-evaluated 45 years ago by Harlan and DeWet (1975), who pointed out that the vast majority of hybridization events rely on meiotic, rather than mitotic, mechanisms, i.e., unreduced gametes rather than mitotic errors. This viewpoint has been reinforced in the intervening years (Ramsey and Schemske, 1998; De Storme and Geelen, 2013; De Storme and Mason, 2014; Mason and Pires, 2015).

Irrespective of the mode of formation, the terms "allopolyploidy" and "autopolyploidy" clearly represent two ends of a cytogenetic and taxonomic conceptual continuum with broadly overlapping suites of characteristic features (Wendel and Doyle, 2005; Carputo et al., 2006). In recent years, the taxonomic definition for autopolyploidy, as arising within a species, and allopolyploidy, as forming between species, has predominated (Spoelhof et al., 2017b). This may be the most useful definition, despite species concept and classification difficulties, particularly for autopolyploid species (Soltis et al., 2007; Barker et al., 2016). Allopolyploidy events can also vary greatly in the amount of divergence between the progenitor genomes. For example, some interspecific hybridization events that lead to "allopolyploidy" may involve species with subgenomes that are less diverged from each other than "autopolyploid" events arising within a highly polymorphic single species. In rice (*Oryza sativa*), for example, hybridization between the two subspecies *japonica* and *indica* to form novel polyploids results in "genomic shock" and allopolyploid-style gene expression partitioning (Zhao et al., 2018), a phenomenon more normally attributed to allopolyploidy (Grover et al., 2012). By contrast, hybridization between taxonomic species in the *Brassica* "C genome" cytodeme can often lead to fully or partially fertile hybrids with predominantly homologous chromosome pairing during meiosis (Kianian and Quiros, 1992; Bothmer et al., 1995).

In addition to auto- and allo-, Stebbins (1947) proposed a new category of polyploids, known as "segmental" allopolyploids. Stebbins actually used both chromosome behavior and genome structural divergence concepts in his application of the term, as at the time chromosome pairing was thought to rely solely on "structure," rather than sequence homology. He first mentions that "*Cytologically, [segmental allopolyploids] are characterized by the presence of multivalents in varying numbers, so that in meiosis they often resemble autopolyploids more than true allopolyploids*." He later states that "*A segmental allopolyploid may, therefore, be defined as an allopolyploid of which the component genomes bear the majority of their chromosomal segments in common, so that the diploid hybrid from which it is derived has good pairing at meiosis, but in which these genomes differ from each other by a large enough number of chromosomal segments or gene combinations so that free interchange between them is barred by partial or complete sterility on the diploid level.*" This latter idea, that of intermediacy between the archetypal poles of autopolyploidy and allopolyploidy, has often been an unstated assumption in the application of the term since Stebbins' first use, rather than the operational definition that segmental allopolyploids show both multivalent and bivalent formation for some portions of the chromosome complement. Today, however, we would most likely characterize these cases as autopolyploids. As pointed out by Sybenga a quarter century ago (Sybenga, 1996), newly formed polyploids often display multivalents, but established autopolyploids instead are characterized by bivalent formation with random partner choice (tetrasomic inheritance).

This distinction between newly formed vs. evolved is important, as it illustrates the connectedness of the terms auto- and allopolyploidy as well as a temporal dimension. That is, autopolyploids may form between divergent germplasm groups within a species, but later evolve fully disomic inheritance and become "allopolyploid-like" (Bingham, 1980), and not necessarily at homogeneous rates throughout the genome. Thus, it seems important to distinguish mode of formation and evolved meiotic behavior.

MECHANISMS OF GENOME STABILIZATION IN POLYPLOIDS

Newly formed polyploids, whether autopolyploids or allopolyploids, face a major challenge in becoming established, that of regulating meiosis (reviewed by Pelé et al., 2018). Meiosis is a tightly controlled process in all organisms, as fertile progeny must be formed from recombinationally variable gametes. This tight regulation commonly breaks down when two genomes are suddenly present instead of one, i.e., when four copies of homologous chromosomes are present instead of two (reviewed by Cifuentes et al., 2010). Cytologically speaking, there are two strategies by which newly formed polyploids might regulate meiosis: "autopolyploid"-type and "allopolyploid"-type. In "autopolyploid"-type meiotic regulation, crossover number and distribution is stringently regulated: only enough crossovers are permitted so that every two chromosomes will be bound by a single crossover. This promotes strict bivalent formation (with random partner choice, i.e., tetrasomic inheritance) despite the presence of four homologous chromosomes, and is thought to be the most common method for autopolyploid meiotic regulation (Cifuentes et al., 2010; Spoelhof et al., 2017b). This is also the stability mechanism known to occur in *Arabidopsis arenosa* autotetraploids (Lloyd and Bomblies, 2016). Tetravalent formation, that is, when crossovers form between four homologous chromosomes to produce a multivalent, is rarely observed in meiosis of stable autopolyploids, and sometimes not even in synthetic autopolyploids (Sybenga, 1996).

In allopolyploid meiosis, strictly homologous pairing requires some mechanism or mechanisms to discriminate subgenomes (Stebbins, 1950). It has long been noted that newly synthesized allopolyploids suffer a higher degree of irregular chromosomal configurations than do their natural analogs, for example, in cotton (reviewed in Endrizzi et al., 1985), showing that evolutionary enforcement of homologous pairing has been selected over time. Mechanisms leading to this stabilization of homologous pairing are mostly unknown and almost certainly vary among the tens of thousands of allopolyploids that exist in plants (e.g., even B chromosomes have been implicated; Taylor and Evans, 1976). Early insights are beginning to emerge into the spectrum of possible molecular determinants of enforcement of homologous pairing. Prevention of non-homologous pairing between subgenomes in allohexaploid bread wheat is facilitated by the major qualitative effect *Ph1* locus (Sears, 1976; Feldman, 1993; Griffiths et al., 2006; Bhullar et al., 2014), for which the molecular mechanism is still not completely characterized, but which may involve suppression of *CDK2*-like activity to result in chromatin modifications (Greer et al., 2012) as well as the presence of an additional copy of a meiotic *ZIP4* gene (Rey et al., 2017). By contrast, at least eight meiosis genes have been implicated in genomic stabilization of autotetraploid *Arabidopsis arenosa* (Yant et al., 2013), with two of these genes (*ASY1* and *ASY3*) later found to directly reduce multivalent formation and chiasma number, as expected (Morgan et al., 2020). *Triticum* and *Aradidopsis* represent the two best-characterized models for meiotic regulation in polyploids to date (reviewed by Cifuentes et al., 2010; Lloyd and Bomblies, 2016). In recent years, great progress has been made toward understanding the molecular mechanisms underlying regulation of meiosis in polyploids (see reviews by Cifuentes et al., 2010; Grandont et al., 2013; Bomblies et al., 2015, 2016; Lloyd and Bomblies, 2016; Pelé et al., 2018). To date, however, only a few polyploid species and synthetics have been investigated for meiotic stability mechanisms; future investigations across the tree of life are necessary to understand the spectrum of meiotic evolutionary responses to polyploidy and which components might be generalizable.

One question of importance is whether meiotic stabilization following polyploidization is a gradual process, or whether allelic variants present in the diploid progenitors can lead to immediately stable allo- or autopolyploids. In synthetic *Brassica* hybrids, the first generation has clearly been established to be the least stable (Szadkowski et al., 2010), following which meiosis may stabilize over time (Prakash et al., 1999; Gaebelein et al., 2019), putatively due to selection for particular allelic complements conferring higher fertility (Gaebelein and Mason, 2018; Gaebelein et al., 2019). Swaminathan and Sulbha (1959) also found that stability increased over 19 generations of selection in a single genotype of autotetraploid *B. rapa* – a surprising result, because under strict self-pollination the only way for stability to arise would be via *de novo* mutation, chromosome rearrangements or changes in epimethylation, as initial plants would be 100% homozygous. However, it is clear that "complete" stabilization (i.e., complete prevention of homoeologous chromosome pairing) does not occur in *Brassica*: inspection of established *B. napus* has revealed high frequencies of chromosome rearrangements in this young allotetraploid species (Chalhoub et al., 2014; Samans et al., 2017; Mason et al., 2018). These results are similar to those found in very recent (∼80 year old) allopolyploid species in *Tragopogon*, where extensive karyotype variation has been observed, including clear products of homoeologous recombination between the subgenomes (Chester et al., 2012). In *Arabidopsis arenosa* polyploids, a gradual process of generational selection for "adapted" meiosis gene alleles has been proposed, based on selective sweeps between diploid and tetraploid populations (Yant et al., 2013). Interestingly, natural populations of tetraploid *Arabidopsis lyrata* seem to have acquired these *Arabidopsis arenosa* alleles which facilitate meiotic stabilization via interspecific hybridization (Marburger et al., 2019), suggesting a possible shortcut to stabilization.

But are all synthetic and newly formed polyploids in fact meiotically unstable? Although this seems to be a common general trend (see Pelé et al., 2018 for review), there are also examples of immediately stable auto- and allopolyploids. For instance, it seems that kale genotypes of *Brassica oleracea* can be induced to form stable autopolyploids (Jenczewski et al., 2002), despite the fact that most autotetraploids in this species are highly unstable (Howard, 1939; Zdráhalová, 1968). Gupta et al. (2016) also found stable meiotic behavior in a single genotype of *de novo* allohexaploid *Brassica* formed by the cross between *B. carinata* and *B. rapa*, despite the fact that the majority of lines from this cross combination are highly unstable (Tian et al., 2010; Gupta et al., 2016). Some allopolyploid species, such as white

clover, also seem to have clear separation between subgenomes, with no indication of instability following allopolyploidization (Griffiths et al., 2019). Possibly, considerable genetic variation is present within some lineages for the frequency and prevention of non-homologous recombination events, which may occur through different mechanisms such as increased stringency of sequence-homology required for crossover formation, timing of condensation of chromosomes belonging to different genomes, and changes in crossover frequency, targeting and distribution, which may be modulated by many different genomic features and which are only now starting to be elucidated (Zhang et al., 2020; see also Cifuentes et al., 2010 for review). Although only speculation, this would perhaps explain why some plant families (e.g., Brassicaceae) have widely varying karyotypes and chromosome numbers even between closely related species (Schranz et al., 2006), while other families (e.g., Solanaceae) have highly conserved karyotypes and chromosome numbers (Wu and Tanksley, 2010).

HOMOEOLOGOUS EXCHANGES IN ALLOPOLYPLOIDS

Mechanisms of polyploid formation and meiotic regulation are important to consider in all polyploids, as "established" polyploid species can still be prone to meiotic errors resulting from imperfect stabilization processes. In many allopolyploid species, "homoeologous exchanges" (HEs) occur via mispairing between ancestrally related chromosomes belonging to different genomes. These exchanges swap pieces of DNA between the subgenomes, and can lead to deletions, duplications and translocations (**Figure 1**). Not all non-homologous exchanges are homoeologous, that is, occur between related genomic segments that have diverged from a common ancestor. Even small regions of duplicated DNA within a genome are sufficient to induce occasional non-homologous chromosome rearrangements. However, these tend to be heavily suppressed, such that recombination between repetitive sequences (which can easily result in genomic instability) is rare (Putnam et al., 2009). Hence, the vast majority of crossovers between non-homologous chromosomes occur between homoeologous regions, particularly in recent allopolyploids (Nicolas et al., 2012).

Homoeologous exchanges can result in either "balanced" or "reciprocal" translocations ("homoeologous reciprocal translocations" or HRTs) which swap the locations of two homoeologous DNA segments, or "unbalanced" or "homoeologous non-reciprocal" translocations (HNRTs) (effectively duplication/deletion events, **Figure 1**). However, this terminology is misleading and should probably be avoided. "Non-reciprocal" exchanges are of course actually reciprocal in terms of crossover events (see **Figure 1**), such that "non-reciprocal" refers only to the products of the exchange, i.e., whether one piece of DNA has been swapped for another, or whether an additional copy of a DNA sequence has replaced the homoeologous copy in a "duplication-deletion" event. Homoeologous exchanges generally occur via co-opting of the homologous recombination pathway, but where homoeologous chromosome regions (ancestrally related stretches of DNA from different subgenomes) act as the substrate instead of homologous chromosomes (Nicolas et al., 2009). Depending on the structural divergence between the subgenomes, whole chromosomes may be "homoeologous" or syntenic in terms of DNA sequence along their entire length, or single chromosomes may contain many small stretches of DNA that are homoeologous to parts of chromosomes in the other subgenome. In many species, recurrent polyploidy events have resulted in both "primary" and "secondary" homoeology: that is, homoeology between subgenomes resulting from a recent allopolyploid event, and

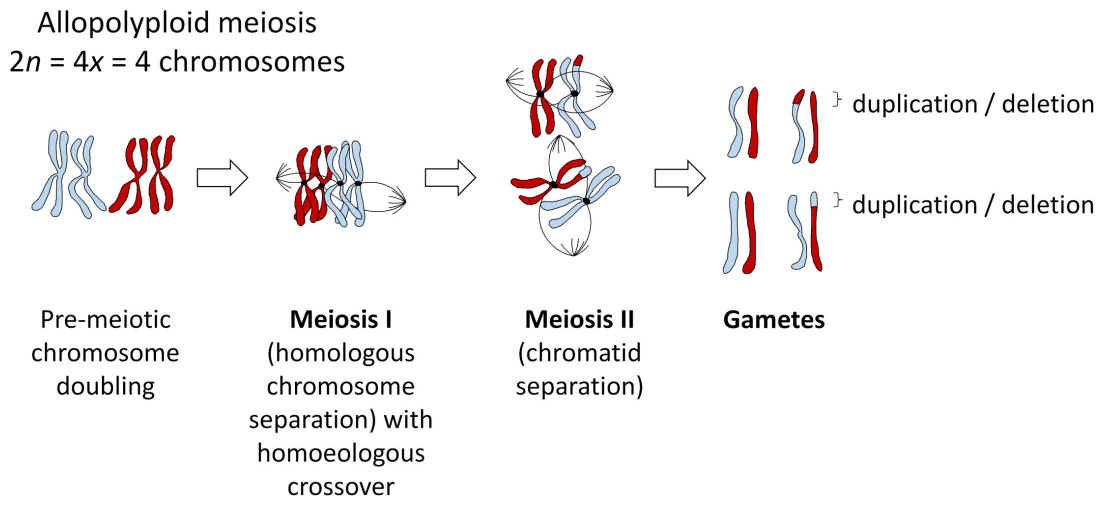


FIGURE 1 | Meiosis in an example allopolyploid with $2n = 4x = 2$ chromosomes (subgenomes indicated in red and blue), showing the most probable outcome of a single crossover event between homoeologous (ancestrally homologous) chromosomes. All such events will most likely be heterozygous in the first generation, even under self-pollination, as gametes with a homoeologous recombination event (duplication/deletion) will unite with gametes from a different meiosis (i.e., pollen with ovules), but may become fixed in subsequent generations after self-pollination.

homoeology within each subgenome between regions resulting from more ancient polyploidy events. A good example of this is provided by the *Brassica* genus: in addition to the three recent allopolyploid species *B. juncea*, *B. napus*, and *B. carinata*, with AABB, AACC, and BBCC genome complements, respectively, for which the A-B, A-C, and B-C genomic relationships, respectively represent primary homoeology, each of the A, B, and C genomes contains triplicated genomic segments resulting from mesopolyploidy events, representing secondary homoeology (Parkin et al., 2003).

Homoeologous exchanges are now known to be common in synthetic polyploids, as well as those of recent evolutionary origin. Recent allopolyploids with commonly detected HEs include *Tragopogon* species (Chester et al., 2012), peanuts (*Arachis hypogaea*; Bertioli et al., 2019; Zhuang et al., 2019), quinoa (*Chenopodium quinoa*; Jarvis et al., 2017), tobacco (*Nicotiana tabacum*; Chen et al., 2018) and rapeseed (*Brassica napus*; Chalhoub et al., 2014). Synthetics with frequent HEs include allopolyploid rice (constructed from *Oryza sativa* subsp. *indica* × subsp. *japonica*; Sun et al., 2017; Li et al., 2019) and *Brassica* species (Song et al., 1995; Gaeta et al., 2007; Szadkowski et al., 2010), as well as many more: this phenomenon may be generalizable across most newly formed polyploids as a result of meiotic instability (reviewed by Pelé et al., 2018).

It should be noted for completeness that homoeologous exchanges are not the only form of genomic instability in novel polyploids. Previously, a great deal of attention has been paid to the activation of transposable elements as a result of "genome shock," a phenomenon first proposed by McClintock (1984). Polyploidy in many species seems to be associated with bursts of transposable element activation (for review see Vicient and Casacuberta, 2017; Nieto-Feliner et al., 2020, this issue). Transposable elements may also (rarely) act as a substrate for non-homologous recombination events (Xiao and Peterson, 2000; Xuan et al., 2012), and also cause sequence mutagenesis after excision due to double strand break repair mechanisms, which often insert or delete a few basepairs during the non-homologous end-joining process (for review see Gorbunova and Levy, 1999). Transposable element activation and novel SSR mutations have been reported in *Brassica* synthetics (Zou et al., 2011; Gao et al., 2014), and widespread loss of non-coding sequences in synthetic wheat polyploids (Ozkan et al., 2001; Shaked et al., 2001), all independent of homoeologous exchange events.

HOMOEOLOGOUS RECOMBINATION EVENTS CAN GENERATE NOVEL VARIATION, AFFECT PHENOTYPE AND ACT AS TARGETS FOR NATURAL SELECTION

Non-homologous recombination events can result in duplications, deletions and chromosome rearrangements. Although karyotypic variation resulting from non-homologous chromosome recombination putatively occurs in almost all evolutionary lineages, facilitating karyotype change over time, it is much more likely that a chromosome rearrangement (particularly a larger deletion or duplication) will prove fatal in a diploid lineage (Schuermann et al., 2005). However, the presence of an extra set of chromosomes provides a "buffer" for chromosome change: when two or more copies of a gene or genomic region are present, this can allow novel variation to arise without impacting viability and fertility to as great an extent. This genomic redundancy, in fact, has classically been considered to be at least partially responsible for the success of polyploidy in many plant lineages (Leitch and Leitch, 2008), although the same redundancy which can buffer high-impact mutations and prevent them from being deleterious may also slow the rate of loss of deleterious alleles and fixation of beneficial alleles (Stebbins, 1971; Otto and Whitton, 2000). Homoeologous recombination events are also, of course, more common in polyploids, which provide millions of potential substrates for non-homologous recombination and formation of crossovers between two similar DNA sequences.

Homoeologous exchanges, as well as presence-absence variants and other karyotypic changes, have now been conclusively linked to phenotypic changes in many species, including polyploid crops (reviewed by Schiessl et al., 2019). In fact, a number of homoeologous exchanges (almost all examples involve duplication-deletion events, as reciprocal translocations are harder to detect) have now been demonstrated to have been selected for in crops: in *Brassica napus* (rapeseed), winter and spring crop types are differentiated by homoeologous exchanges involving major flowering time regulators such as *FLC* (Schiessl et al., 2017), and effects of homoeologous exchanges on disease resistance and glucosinolate metabolism have also been observed (Stein et al., 2017; Hurgobin et al., 2018). In allotetraploid peanut, fixed homoeologous exchanges (duplication-deletion events) were seen to generate phenotypic novelty, with direct effects on flower color (Bertioli et al., 2019). In many synthetic hybrids produced from a single homozygous individual, homoeologous exchanges lead to generation of major genetic and phenotypic novelty (Xiong et al., 2011; Spoelhof et al., 2017a; Sun et al., 2017; Li et al., 2019). Hence, homoeologous exchanges (both duplication-deletions and chromosome rearrangements) may comprise an important evolutionary substrate for divergence, speciation and adaptation in newly formed allopolyploids.

It is of interest to consider the possible relationships between HEs and the constraints on genic retention and evolution following whole genome doubling imposed by selection at the gene balance level (Birchler and Veitia, 2010, 2012, 2014). To the extent that gene content and function are equivalent among homoeologous segments, HEs would not, to a first approximation, appear to materially impact gene balance. In the case of allopolyploids, however, where there almost certainly are both functional and copy-number differences among homoeologs, it seems likely that the selective fate or survivorship of particular HEs might in part be directed by gene-balance considerations. As this is an entirely unexplored relationship, it represents a natural area for future investigation.

HOMOEOLOGOUS EXCHANGES AND SEGMENTAL ALLOPOLYPLOIDY

Reconciliation between the classic definition of segmental allopolyploids as "containing autopolyploid and allopolyploid segments" and modern genomic observations of genic synteny is, of course, provided by the mechanism of homoeologous exchange following allopolyploidy. A modern conception of "segmental allopolyploids" may thus include both transitional autopolyploids as well as allopolyploids that contain a mix of auto- and allopolyploid segments derived via homoeologous exchanges (duplication-deletion events; e.g., Sun et al., 2017; Leal-Bertioli et al., 2018; Bertioli et al., 2019). The dynamics of this process have now been described in numerous experimental systems.

In homoeologous exchanges, the products of a single event can be described as either "balanced" exchanges, or as "duplication-deletion" events, where the latter are hypothetically more common due to random segregation of chromatids following a homoeologous crossover event. In these "duplication-deletion" events, segments of one subgenome are deleted and replaced by segments of the other subgenome. If these events become fixed, then this genomic region is in fact "autopolyploid," with four copies of the same subgenome, while the rest of the genome remains allopolyploid. If no selection or bias is present, this would generate a complex mosaic of genomic regions representing one or the other subgenome, or both (**Figure 2**), with considerable relevance to phylogenomics and phylogenetics (Edger et al., 2018). An excellent recent example of this process is provided by genomic investigations of allotetraploid peanut, in which regions of both A and B subgenomes had been replaced by copies of the other subgenome (AABB –> AAAA or BBBB) (Bertioli et al., 2019), putatively as a result of fixation of these homoeologous exchanges after a single allopolyploidization event. On the other hand, a possible outcome of biased replacement of one subgenome with the other subgenome, as has been documented to occur via fertility-based selection in some species (e.g., Gaebelein et al., 2019), could make an even more interesting pattern: an "autopolyploid" may result, but possibly one that appears to have small genomic regions introgressed from another species (**Figure 2**). Recently, synthetic rice polyploids formed by hybridization between *japonica* and *indica* subspecies also revealed directional loss of one subgenome through selection for the products of HEs (Zhang et al., 2019). The authors found that this "homogenization" (retention of two copies of one subgenome and loss of the corresponding homoeologous copy from the other subgenome) also altered gene expression and enhanced alternative splicing in these chromosome regions, thus suggesting a possible selective mechanism for these events. Although an interesting speculation,

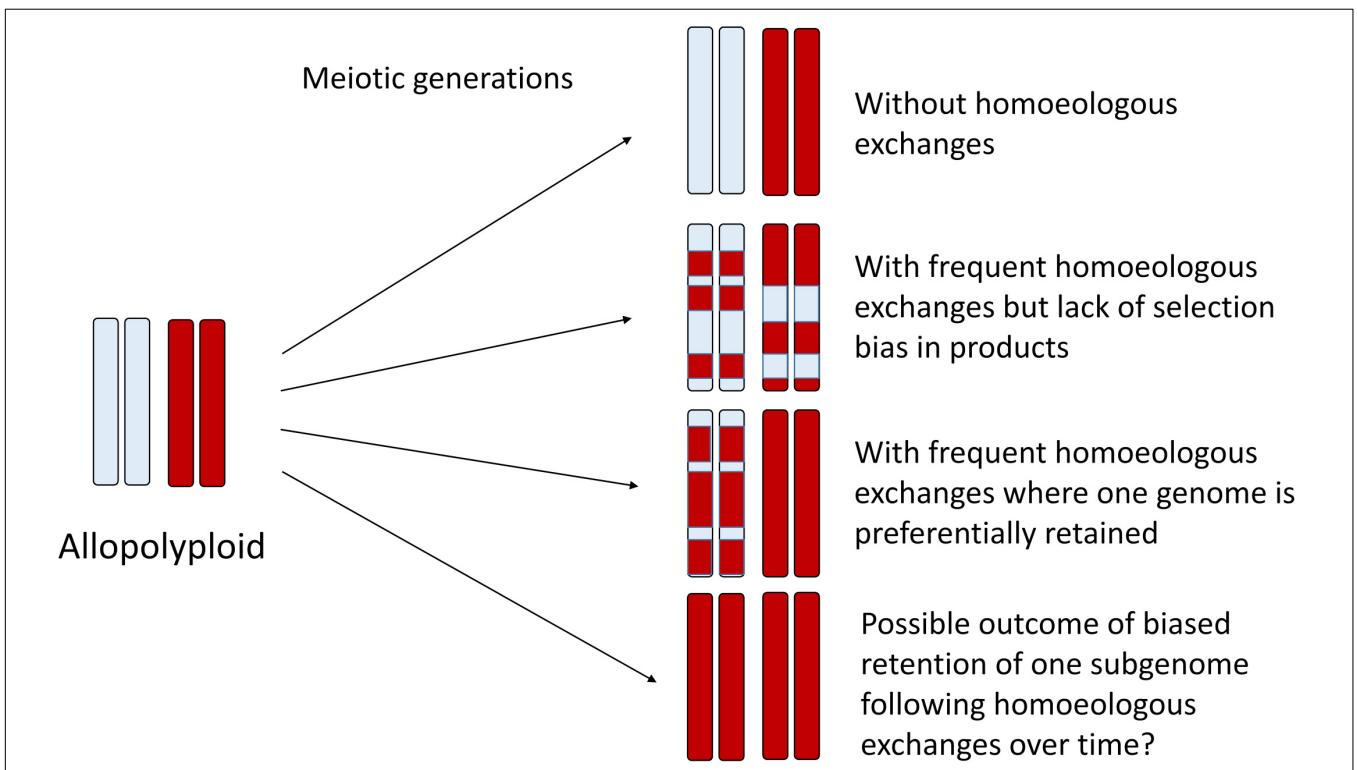


FIGURE 2 | Homoeologous exchanges can generate a diverse spectrum of genomic mosaics over the generations, where some regions of the genome retain homoeologous segments and others become genomically homozygous for a single parental homoeolog, as illustrated here for one pair of homoeologous chromosomes. Thus, some regions of the genome might appear to be "autopolyploid" whereas others appear "allopolyploid." At the population level and over time, HEs may generate highly variable progeny that may be subject to natural selection, thus fixing specific chromosomal recombinants. In the limit, directional selection may favor one progenitor homoeolog, which may thus appear to have an autopolyploid origin. Genic divergence for duplicates is expected to reflect this history.

a uni-directional process has so far only been observed in very recent and synthetic allopolyploids. Under natural conditions, negative selective pressure against *de novo* HEs would also have to be overcome. Minority cytotype disadvantage, where individuals heterozygous for particular chromosome rearrangements have lower reproductive success, could play a role in purging novel homoeologous exchanges from populations, and likely does in most cases. Finally, it is unclear the extent to which initial conditions established at the time of allopolyploid formation, that is, the genomic features of the progenitor diploids, are determinative of the future survivorship of HEs in their derived allopolyploids. More specifically, do the same, still somewhat mysterious genomic features that are thought to be involved in the establishment of subgenome dominance (Cheng et al., 2018; Wendel et al., 2018) play an important role in the genomic distribution and selective fate of HEs? This too represents an area for future research.

DISCUSSION

Here we have attempted to provide a synopsis of our growing recognition that homoeologous exchange following polyploidy is a common evolutionary process leading to genomically variable progeny that can serve as substrates for natural selection. Thus, HEs comprise an important dimension of polyploid genomics, potentially representing a key mechanism of post-polyploidization diversification and speciation. This same process has implications for the inference of polyploid parentage and our understanding of the terms autopolyploidy and allopolyploidy, as well as "segmental allopolyploidy." In this respect, phylogenetic or phylogenomic analyses will benefit from consideration of the genomic mosaicism potentially generated by HEs (Edger et al., 2018), and the possibility of conversion of a strict allopolyploid to a partially autopolyploid genome through homoeologous exchanges. The use of synthetic systems, where historical polyploidization events are "recreated" by crossing between diploid progenitor species, may help shed light on the spectrum of mechanisms and outcomes involved in the early stages of allopolyploid genome evolution (Xiong et al., 2011; Samans et al., 2017; Spoelhof et al., 2017a; Sun et al., 2017; Li et al., 2019), as may investigation of very young allopolyploids such as *Tragopogon mirus* and *T. miscellus* (Buggs et al., 2011, 2012), *Senecio cambrensis* (Ashton and Abbott, 1992; Hegarty et al., 2006), *Mimulus peregrinus* (Vallejo-Marín et al., 2015) and *Spartina anglica* (Baumel et al., 2002; Ainouche et al., 2004). In particular, better understanding of the mechanisms controlling genome stability (i.e., frequency of non-homologous recombination events and other mutations) and the possible genotypic influences on these mechanisms may prove a fruitful avenue for further investigation. Homoeologous exchanges in allopolyploids in particular may have far-reaching implications for polyploid evolution, providing evolutionary novelty, helping stabilize genomes and facilitating speciation. Our appreciation of the significance of HEs in polyploid evolution will undoubtedly be enhanced by the increasing application of genomic tools to natural (Bomblies et al., 2015; Yant and Bomblies, 2017; Marburger et al., 2019) and synthetic (Samans et al., 2017; Sun et al., 2017; Hurgobin et al., 2018; Li et al., 2019) polyploid complexes, combined with an increasing experimental focus on cytogenetic and meiotic mechanisms.

AUTHOR CONTRIBUTIONS

JW and AM conceptualized and wrote the manuscript. AM produced the figures with critical input from JW. Both authors contributed to the article and approved the submitted version.

FUNDING

AM acknowledges funding support from Deutsche Forschungsgemeinschaft grants MA6473/1-1, MA6473/2-1, and DFG Sino-German grant MA6473/3-1. JW acknowledges the support of the US National Science Foundation and Cotton Incorporated.

REFERENCES

Abel, S., and Becker, H. C. (2007). The effect of autopolyploidy on biomass production in homozygous lines of *Brassica rapa* and *Brassica oleracea*. *Plant Breed.* 126, 642–643. doi: 10.1111/j.1439-0523.2007.01405.x

Ainouche, M. L., Baumel, A., and Salmon, A. (2004). *Spartina anglica* C. E. Hubbard: a natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biol. J. Linn. Soc.* 82, 475–484. doi: 10.1111/j.1095-8312.2004.00334.x

Ashton, P. A., and Abbott, R. J. (1992). Multiple origins and genetic diversity in the newly arisen allopolyploid species, *Senecio cambrensis* Rosser (Compositae). *Heredity* 68, 25–32. doi: 10.1038/hdy.1992.3

Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., and Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210, 391–398. doi: 10.1111/nph.13698

Baumel, A., Ainouche, M., Kalendar, R., and Schulman, A. H. (2002). Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). *Mol. Biol. Evol.* 19, 1218–1227. doi: 10.1093/oxfordjournals.molbev.a004182

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., et al. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* 51, 877–884. doi: 10.1038/s41588-019-0405-z

Bhullar, R., Nagarajan, R., Bennypaul, H., Sidhu, G. K., Sidhu, G. K., Rustgi, S., et al. (2014). Silencing of a metaphase I-specific gene results in a phenotype similar to that of the *Pairing homeologous 1* (*Ph1*) gene mutations. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14187–14192. doi: 10.1073/pnas.1416241111

Bingham, E. T. (1980). "Maximising heterozygosity in autopolyploids," in *Polyploidy: Biological Relevance*, ed. W. H. Lewis (London: Plenum Press), 471–489. doi: 10.1007/978-1-4613-3069-1_24

Birchler, J. A., and Veitia, R. A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 186, 54–62. doi: 10.1111/j.1469-8137.2009.03087.x

Birchler, J. A., and Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14746–14753. doi: 10.1073/pnas.1207726109

Birchler, J. A., and Veitia, R. A. (2014). "The gene balance hypothesis: dosage effects in plants," in *Plant Epigenetics and Epigenomics*, eds C. Spillane and

P. C. McKeown (Totowa, NJ: Humana Press), 25–32. doi: 10.1007/978-1-62703-773-0_2

Bomblies, K., Higgins, J. D., and Yant, L. (2015). Meiosis evolves: adaptation to external and internal environments. *New Phytol.* 208, 306–323. doi: 10.1111/nph.13499

Bomblies, K., Jones, G., Franklin, C., Zickler, D., and Kleckner, N. (2016). The challenge of evolving stable polyploidy: Could an increase in "crossover interference distance" play a central role? *Chromosoma* 125, 287–300. doi: 10.1007/s00412-015-0571-4

Bothmer, R., Gustafsson, M., and Snogerup, S. (1995). *Brassica* sect. *Brassica (Brassicaceae)* II. Inter- and intraspecific crosses with cultivars of *B. oleracea*. *Genet. Resour. Crop Evol.* 42, 165–178. doi: 10.1007/bf02539520

Buggs, R. J. A., Renny-Byfield, S., Chester, M., Jordon-Thaden, I. E., Viccini, L. F., Chamala, S., et al. (2012). Next-generation sequencing and genome evolution in allopolyploids. *Am. J. Bot.* 99, 372–382. doi: 10.3732/Ajb.1100395

Buggs, R. J. A., Zhang, L., Miles, N., Tate, J. A., Gao, L., Wei, W., et al. (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* 21, 551–556. doi: 10.1016/j.cub.2011.02.016

Carputo, D., Camadro, E. L., and Peloquin, S. J. (2006). Terminology for polyploids based on cytogenetic behavior: consequences in genetics and breeding. *Plant Breed. Rev.* 26, 105–124. doi: 10.1002/9780470650325.ch4

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chen, S., Ren, F., Zhang, L., Liu, Y., Chen, X., Li, Y., et al. (2018). Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. *Mol. Plant.* 11, 914–927. doi: 10.1016/j.molp.2018.04.009

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4, 258–268. doi: 10.1038/s41477-018-0136-7

Chester, M., Gallagher, J. P., Symonds, V. V., da Silva, A. V. C., Mavrodiev, E. V., Leitch, A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. U.S.A.* 109, 1176–1181. doi: 10.1073/pnas.1112041109

Cifuentes, M., Grandont, L., Moore, G., Chèvre, A. M., and Jenczewski, E. (2010). Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol.* 186, 29–36. doi: 10.1111/j.1469-8137.2009.03084.x

De Storme, N., and Geelen, D. (2013). Sexual polyploidization in plants–cytological mechanisms and molecular regulation. *New Phytol.* 198, 670–684. doi: 10.1111/nph.12184

De Storme, N., and Mason, A. S. (2014). Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* 1, 10–33. doi: 10.1016/j.cpb.2014.09.002

Edger, P. P., McKain, M. R., Bird, K. A., and VanBuren, R. (2018). Subgenome assignment in allopolyploids: challenges and future directions. *Curr. Opin. Plant Biol.* 42, 76–80. doi: 10.1016/j.pbi.2018.03.006

Endrizzi, J. E., Turcotte, E. L., and Kohel, R. J. (1985). Genetics, cytology, and evolution of *Gossypium*. *Adv. Genet.* 23, 271–375. doi: 10.1016/s0065-2660(08)60515-5

Feldman, M. (1993). Cytogenetic activity and mode of action of the pairing homoeologous (*Ph1*) gene of wheat. *Crop Sci.* 33, 894–897. doi: 10.2135/cropsci1993.0011183x003300050003x

Gaebelein, R., and Mason, A. S. (2018). Allohexaploids in the genus *Brassica*. *Crit. Rev. Plant Sci.* 37, 422–437. doi: 10.1080/07352689.2018.1517143

Gaebelein, R., Schiessl, S. V., Samans, B., Batley, J., and Mason, A. S. (2019). Inherited allelic variants and novel karyotype changes influence fertility and genome stability in *Brassica* allohexaploids. *New Phytol.* 223, 965–978. doi: 10.1111/nph.15804

Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., and Osborn, T. C. (2007). Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19, 3403–3417. doi: 10.1105/tpc.107.054346

Gao, C., Yin, J., Mason, A. S., Tang, Z., Ren, X., Li, C., et al. (2014). Regularities in simple sequence repeat variations induced by a cross of resynthesized *Brassica napus* and natural *Brassica napus*. *Plant OMICS* 7, 35–46.

Gorbunova, V., and Levy, A. A. (1999). How plants make ends meet: DNA double-strand break repair. *Trends Plant Sci.* 4, 263–269. doi: 10.1016/S1360-1385(99)01430-2

Grandont, L., Jenczewski, E., and Lloyd, A. (2013). Meiosis and its deviations in polyploid plants. *Cytogenet. Genome Res.* 140, 171–184. doi: 10.1159/000351730

Greer, E., Martin, A. C., Pendle, A., Colas, I., Jones, A. M. E., Moore, G., et al. (2012). The *Ph1* locus suppresses *Cdk2*-type activity during premeiosis and meiosis in wheat. *Plant Cell* 24, 152–162. doi: 10.1105/tpc.111.094771

Griffiths, A. G., Moraga, R., Tausen, M., Gupta, V., Bilton, T. P., Campbell, M. A., et al. (2019). Breaking free: the genomics of allopolyploidy-facilitated niche expansion in white clover. *Plant Cell* 31, 1466–1487. doi: 10.1105/tpc.18.00606

Griffiths, S., Sharp, R., Foote, T. N., Bertin, I., Wanous, M., Reader, S., et al. (2006). Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature* 439, 749–752. doi: 10.1038/nature04434

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Gupta, M., Atri, C., Agarwal, N., and Banga, S. S. (2016). Development and molecular-genetic characterization of a stable *Brassica* allohexaploid. *Theor. Appl. Genet.* 129, 2085–2100. doi: 10.1007/s00122-016-2759-2

Harlan, J. R., and DeWet, J. M. J. (1975). On Ö. Winge and a prayer: the origins of polyploidy. *Bot. Rev.* 41, 361–390. doi: 10.1007/bf02860830

Hegarty, M. J., Barker, G. L., Wilson, I. D., Abbott, R. J., Edwards, K. J., and Hiscock, S. J. (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr. Biol.* 16, 1652–1659. doi: 10.1016/j.cub.2006.06.071

Howard, H. W. (1939). The cytology of autotetraploid kale, *Brassica oleracea*. *Cytologia* 10, 77–87. doi: 10.1508/cytologia.10.77

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi: 10.1038/nature21370

Jenczewski, E., Eber, F., Manzanares-Dauleux, M. J., Chèvre, A. M., and Chevre, A. M. (2002). A strict diploid-like pairing regime is associated with tetrasomic segregation in induced autotetraploids of kale. *Plant Breed.* 121, 177–179. doi: 10.1046/j.1439-0523.2002.00672.x

Kianian, S. F., and Quiros, C. F. (1992). Trait inheritance, fertility, and genomic relationships of some n = 9 *Brassica* species. *Genet. Resour. Crop Evol.* 39, 165–175. doi: 10.1007/BF00051930

Kihara, H., and Ono, T. (1926). Chromosomenzahlen und systematische Gruppierung der *Rumex*-Arten. *Z. Zellforsch. Microsk. Anat.* 4, 475–481. doi: 10.1007/bf00391215

Leal-Bertioli, S. C. M., Godoy, I. J., Santos, J. F., Doyle, J. J., Guimarães, P. M., Abernathy, B. L., et al. (2018). Segmental allopolyploidy in action: Increasing diversity through polyploid hybridization and homoeologous recombination. *Am. J. Bot.* 105, 1053–1066. doi: 10.1002/ajb2.1112

Leitch, A. R., and Leitch, I. J. (2008). Genomic plasticity and the diversity of polyploid plants. *Science* 320, 481–483. doi: 10.1126/science.1153585

Li, N., Xu, C., Zhang, A., Lv, R., Meng, X., Lin, X., et al. (2019). DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytol.* 223, 979–992. doi: 10.1111/nph.15820

Lloyd, A., and Bomblies, K. (2016). Meiosis in autopolyploid and allopolyploid *Arabidopsis*. *Curr. Opin. Plant Biol.* 30, 116–122. doi: 10.1016/j.pbi.2016.02.004

Marburger, S., Monnahan, P., Seear, P. J., Martin, S. H., Koch, J., Paajanen, P., et al. (2019). Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* 10:5218. doi: 10.1038/s41467-019-13159-5

Mason, A. S., Chauhan, P., Banga, S., Banga, S. S., Salisbury, P., Barbetti, M. J., et al. (2018). Agricultural selection and presence-absence variation in spring-type canola germplasm. *Crop Pasture Sci.* 69, 55–64. doi: 10.1071/CP17161

Mason, A. S. A. S., and Pires, J. C. C. (2015). Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet.* 31, 5–10. doi: 10.1016/j.tig.2014.09.011

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

Morgan, C., Zhang, H., Henry, C. E., Franklin, C. F. H., and Bomblies, K. (2020). Derived alleles of two axis proteins affect meiotic traits in autotetraploid *Arabidopsis arenosa*. *Proc. Natl. Acad. Sci. U.S.A.* 117, 8980–8988. doi: 10.1073/pnas.1919459117

Nicolas, S. D., Leflon, M., Monod, H., Eber, F., Coriton, O., Huteau, V., et al. (2009). Genetic regulation of meiotic cross-overs between related genomes in *Brassica napus* haploids and hybrids. *Plant Cell* 21, 373–385. doi: 10.1105/tpc.108.062273

Nicolas, S. D., Monod, H., Eber, F., Chèvre, A. M., and Jenczewski, E. (2012). Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. *Plant J.* 70, 691–703. doi: 10.1111/j.1365-313X.2012.04914.x

Nieto-Feliner, G., Casacuberta, J., and Wendel, J. F. (2020). Genomics of evolutionary novelty in hybrids and polyploids. *Front. Genet.* 11:792. doi: 10.3389/fgene.2020.00792

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401

Ozkan, H., Levy, A. A., and Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13, 1735–1747. doi: 10.1105/tpc.13.8.1735

Parkin, I. A. P., Sharpe, A. G., and Lydiate, D. J. (2003). Patterns of genome duplication within the Brassica napus genome. *Genome* 46, 291–303. doi: 10.1139/G03-006

Pelé, A., Rousseau-Gueutin, M., and Chèvre, A.-M. (2018). Speciation success of polyploid plants closely relates to the regulation of meiotic recombination. *Front. Plant Sci.* 9:907. doi: 10.3389/fpls.2018.00907

Prakash, S., Takahata, Y., Kirti, P. B., and Chopra, V. L. (1999). "Cytogenetics," in *Biology of Brassica coenospecies*, ed. C. Gómez-Campo (Amsterdam: Elsevier Science B.V.), 59–106.

Putnam, C. D., Hayes, T. K., and Kolodner, R. D. (2009). Specific pathways prevent duplication-mediated genome rearrangements. *Nature* 460, 984–989. doi: 10.1038/nature08217

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Rey, M. D., Martín, A. C., Higgins, J., Swarbreck, D., Uauy, C., Shaw, P., et al. (2017). Exploiting the ZIP4 homologue within the wheat Ph1 locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. *Mol. Breed.* 37:95. doi: 10.1007/s11032-017-0700-2

Samans, B., Chalhoub, B., and Snowdon, R. J. (2017). Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. *Plant Genome* 10, 1–15. doi: 10.3835/plantgenome2017.02.0013

Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. (2017). Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci. Rep.* 7:41485. doi: 10.1038/srep41845

Schiessl, S. V., Katche, E., Ihien, E., Chawla, H. S., and Mason, A. S. (2019). The role of genomic structural variation in the genetic improvement of polyploid crops. *Crop J.* 7, 127–140. doi: 10.1016/j.cj.2018.07.006

Schranz, M. E., Lysak, M. A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* 11, 535–542. doi: 10.1016/j.tplants.2006.09.002

Schuermann, D., Molinier, J., Fritsch, O., and Hohn, B. (2005). The dual nature of homologous recombination in plants. *Trends Genet.* 21, 172–181. doi: 10.1016/j.tig.2005.01.002

Sears, E. R. (1976). Genetic control of chromosome pairing in wheat. *Annu. Rev. Genet.* 10, 31–51. doi: 10.1146/annurev.ge.10.120176.000335

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13, 1749–1759. doi: 10.1105/tpc.010083

Soltis, D. E., Soltis, P. S., Schemske, D. W., Hancock, J. F., Thompson, J. N., Husband, B. C., et al. (2007). Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56, 13–30. doi: 10.2307/25065732

Soltis, D. E., Visger, C. J., and Soltis, P. S. (2014). The polyploidy revolution then…and now: stebbins revisited. *Am. J. Bot.* 101, 1057–1078. doi: 10.3732/ajb.1400178

Song, K. M., Lu, P., Tang, K. L., and Osborn, T. C. (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7719–7723. doi: 10.1073/pnas.92.17.7719

Spoelhof, J. P., Chester, M., Rodriguez, R., Geraci, B., Heo, K., Mavrodiev, E., et al. (2017a). Karyotypic variation and pollen stainability in resynthesized allopolyploids *Tragopogon miscellus* and *T. mirus*. *Am. J. Bot.* 104, 1484–1492. doi: 10.3732/ajb.1700180

Spoelhof, J. P., Soltis, P. S., and Soltis, D. E. (2017b). Pure polyploidy: closing the gaps in autopolyploid research. *J. Syst. Evol.* 55, 340–352. doi: 10.1111/jse.12253

Stebbins, G. L. (1947). Types of polyploids: their classification and significance. *Adv. Genet.* 1, 403–429. doi: 10.1016/S0065-2660(08)60490-3

Stebbins, G. L. (1950). *Variation and Evolution in Plants*. London: Oxford University Press.

Stebbins, G. L. (1971). *Chromosomal Evolution in Higher Plants*. London: Edward Arnold.

Stein, A., Coriton, O., Rousseau-Gueutin, M., Samans, B., Schiessl, S. V., Obermeier, C., et al. (2017). Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant Biotechnol. J.* 15, 1478–1489. doi: 10.1111/pbi.12732

Sun, Y., Wu, Y., Yang, C., Sun, S., Lin, X., Liu, L., et al. (2017). Segmental allotetraploidy generates extensive homoeologous expression rewiring and phenotypic diversity at the population level in rice. *Mol. Ecol.* 26, 5451–5466. doi: 10.1111/mec.14297

Swaminathan, M. S., and Sulbha, K. (1959). Multivalent frequency and seed fertility in raw and evolved tetraploids of *Brassica campestris* var. *toria*. *Z. Vererbungsl.* 90, 385–392. doi: 10.1007/BF00888813

Sybenga, J. (1996). Chromosome pairing affinity and quadrivalent formation in polyploids: Do segmental allopolyploids exist? *Genome* 39, 1176–1184. doi: 10.1139/g96-148

Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Huneau, C., Belcram, H., et al. (2010). The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol.* 186, 102–112. doi: 10.1111/j.1469-8137.2010.03182.x

Taylor, I. B., and Evans, G. M. (1976). The effect of B chromosomes on homoeologous pairing in species hybrids. *Chromosoma* 57, 25–32. doi: 10.1007/bf00292948

Tian, E., Jiang, Y., Chen, L., Zou, J., Liu, F., and Meng, J. (2010). Synthesis of a *Brassica* trigenomic allohexaploid (*B. carinata* × *B. rapa) de novo* and its stability in subsequent generations. *Theor. Appl. Genet.* 121, 1431–1440. doi: 10.1007/s00122-010-1399-1

Vallejo-Marín, M., Buggs, R. J. A., Cooley, A. M., and Puzey, J. R. (2015). Speciation by genome duplication: repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. *Evolution* 69, 1487–1500. doi: 10.1111/evo.12678

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

Vicient, C. M., and Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* 120, 195–207. doi: 10.1093/aob/mcx078

Wendel, J., and Doyle, J. (2005). "Polyploidy and evolution in plants," in *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, ed. R. J. Henry (Wallingford: CABI Publishing), 97–117. doi: 10.1079/9780851999043.0097

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Wu, F., and Tanksley, S. D. (2010). Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 11:182. doi: 10.1186/1471-2164-11-182

Xiao, Y. L., and Peterson, T. (2000). Intrachromosomal homologous recombination in *Arabidopsis* induced by a maize transposon. *Mol. Gen. Genet.* 263, 22–29. doi: 10.1007/PL00008672

Xiong, Z. Y., Gaeta, R. T., and Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7908–7913. doi: 10.1073/pnas.1014138108

Xuan, Y. H., Zhang, J., Peterson, T., and Han, C. (2012). *Ac*/*Ds* -induced chromosomal rearrangements in rice genomes. *Mob. Genet. Elements* 2, 67–71. doi: 10.4161/mge.20264

Yant, L., and Bomblies, K. (2017). Genomic studies of adaptive evolution in outcrossing Arabidopsis species. *Curr. Opin. Plant Biol.* 36, 9–14. doi: 10.1016/j.pbi.2016.11.018

Yant, L., Hollister, J. D., Wright, K. M., Arnold, B. J., Higgins, J. D., Franklin, F. C. H., et al. (2013). Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* 23, 2151–2156. doi: 10.1016/j.cub.2013.08.059

Yuan, Y., Bayer, P. E., Batley, J., and Edwards, D. (2017). Improvements in genomic technologies: application to crop genomics. *Trends Biotechnol.* 35, 547–558. doi: 10.1016/j.tibtech.2017.02.009

Zdráhalová, B. (1968). Frequency of polyvalents in relation to fertility of autogamized autotetraploid kohlrabi (*Brassica oleracea* var. *gongylodes* L.), cabbage (*Brassica oleracea* var. *capitata* L.) and their hybrids. *Biol. Plant.* 10, 435–444. doi: 10.1007/bf02920986

Zhang, Z., Fu, T., Liu, Z., Wang, X., Xun, H., Li, G., et al. (2019). Extensive changes in gene expression and alternative splicing due to homoeologous exchange in rice segmental allopolyploids. *Theor. Appl. Genet.* 132, 2295–2308. doi: 10.1007/s00122-019-03355-8

Zhang, Z., Gou, X., Xun, H., Bian, Y., Ma, X., Li, J., et al. (2020). Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proc. Natl. Acad. Sci. U.S.A.* 117, 14561–14571. doi: 10.1073/pnas.2003505117

Zhao, L., Han, L., Xiao, C., Lin, X., Xu, C., and Yang, C. (2018). Rapid and pervasive development- and tissue-specific homeolog expression partitioning in newly formed inter-subspecific rice segmental allotetraploids. *BMC Genomics* 19:756. doi: 10.1186/s12864-018-5150-7

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876. doi: 10.1038/s41588-019-0402-2

Zou, J., Fu, D., Gong, H., Qian, W., Xia, W., Pires, J. C., et al. (2011). De novo genetic variation associated with retrotransposon activation, genomic rearrangements and trait variation in a recombinant inbred line population of *Brassica napus* derived from interspecific hybridization with *Brassica rapa*. *Plant J.* 68, 212–224. doi: 10.1111/j.1365-313X.2011.04679.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 September 2020
doi: 10.3389/fgene.2020.565854

Experimental and Field Data Support Range Expansion in an Allopolyploid Arabidopsis Owing to Parental Legacy of Heavy Metal Hyperaccumulation

Timothy Paape[1]†, Reiko Akiyama[1], Teo Cereghetti[1†], Yoshihiko Onda[2], Akira S. Hirao[2,3], Tanaka Kenta[2] and Kentaro K. Shimizu[1,4]**

[1] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [2] Sugadaira Montane Research Center, University of Tsukuba, Tsukuba, Japan, [3] Faculty of Symbiotic Systems Science, Fukushima University, Fukushima, Japan, [4] Kihara Institute for Biological Research, Yokohama City University, Yokohama, Japan

Empirical evidence is limited on whether allopolyploid species combine or merge parental adaptations to broaden habitats. The allopolyploid *Arabidopsis kamchatica* is a hybrid of the two diploid parents *Arabidopsis halleri* and *Arabidopsis lyrata*. *A. halleri* is a facultative heavy metal hyperaccumulator, and may be found in cadmium (Cd) and zinc (Zn) contaminated environments, as well as non-contaminated environments. *A. lyrata* is considered non-tolerant to these metals, but can be found in serpentine habitats. Therefore, the parents have adaptation to different environments. Here, we measured heavy metals in soils from native populations of *A. kamchatica.* We found that soil Zn concentration of nearly half of the sampled 40 sites was higher than the critical toxicity level. Many of the sites were near human construction, suggesting adaptation of *A. kamchatica* to artificially contaminated soils. Over half of the *A. kamchatica* populations had $>1{,}000\ \mu g\ g^{-1}$ Zn in leaf tissues. Using hydroponic treatments, most genotypes accumulated $>3{,}000\ \mu g\ g^{-1}$ Zn, with high variability among them, indicating substantial genetic variation in heavy metal accumulation. Genes involved in heavy metal hyperaccumulation showed an expression bias in the *A. halleri*-derived homeolog in widely distributed plant genotypes. We also found that two populations were found growing on serpentine soils. These data suggest that *A. kamchatica* can inhabit a range of both natural and artificial soil environments with high levels of ions that either of the parents specializes and that it can accumulate varying amount of heavy metals. Our field and experimental data provide a compelling example of combining genetic toolkits for soil adaptations to expand the habitat of an allopolyploid species.

Keywords: adaptation, expression ratio, heavy metal hyperaccumulation, homeolog, polyploid speciation, quantitative variation

OPEN ACCESS

Edited by:
Jonathan F. Wendel,
Iowa State University, United States

Reviewed by:
Roswitha Schmickl,
Academy of Sciences of the Czech Republic (ASCR), Czechia
Jonathan M. Flowers,
New York University Abu Dhabi, United Arab Emirates

****Correspondence:***
Timothy Paape
tpaape@bnl.gov
Kentaro K. Shimizu
kentaro.shimizu@ieu.uzh.ch

†Present address:
Timothy Paape,
Brookhaven National Laboratory, Biology Department, Upton, NY, United States
Teo Cereghetti,
Institute of Integrative Biology, ETH Zürich, Zurich, Switzerland; Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *26 May 2020*
Accepted: *04 September 2020*
Published: *30 September 2020*

Citation:
Paape T, Akiyama R, Cereghetti T, Onda Y, Hirao AS, Kenta T and Shimizu KK (2020) Experimental and Field Data Support Range Expansion in an Allopolyploid Arabidopsis Owing to Parental Legacy of Heavy Metal Hyperaccumulation. Front. Genet. 11:565854. doi: 10.3389/fgene.2020.565854

INTRODUCTION

The ecological advantages of whole-genome duplication and its implication for species habitats have been debated for decades (Ohno, 1970; Stebbins, 1971; Comai, 2005; Soltis et al., 2014). Allopolyploids are interspecific hybrids of two or more parental species that inherit unreduced sets of chromosomes from their parental species. Allopolyploid species may have the potential to combine or merge parental adaptations (Doyle et al., 2008; Buggs et al., 2014; Huynh et al., 2020), which may provide enhanced abilities to tolerate extreme conditions (Levin, 2002; Adams, 2007). Inherited adaptations to abiotic conditions may also provide allopolyploids with the ability to inhabit areas beyond those of the parental species (Shimizu-Inatsugi et al., 2016; Van de Peer et al., 2017; Akiyama et al., 2019; Sun et al., 2020). The evidence of range expansion following allopolyploidization is largely circumstantial (Hoffmann, 2005; Blaine Marchant et al., 2016), therefore identifying ecologically relevant traits that may contribute to the broadening of habitats by allopolyploid species is essential (Ramsey and Ramsey, 2014).

The allotetraploid species *Arabidopsis kamchatica* has one of the broadest distributions of any *Arabidopsis* species (Hoffmann, 2005; Shimizu-Inatsugi et al., 2009) and offers a unique system to examine environmental responses at phenotypic and molecular levels. The species is a natural hybrid of the diploid parents *Arabidopsis halleri*, a hyperaccumulator of the heavy metals zinc (Zn) and cadmium (Cd), and *Arabidopsis lyrata*, a non-hyperaccumulator but with serpentine adaptation. Hyperaccumulating plant species, such as *A. halleri*, transport large amounts of toxic heavy metals (e.g., Zn and Cd) from the roots to the aerial parts of the plant and are also hyper-tolerant to high concentrations of these metals in soils (Bert et al., 2000; Pauwels et al., 2006; Krämer, 2010), whereas transport of heavy metals from roots to shoots is prevented in non-hyperaccumulators such as *A. lyrata* (Paape et al., 2016).

Metal concentrations in soils are an important factor for environmental niches, because high concentrations are toxic for most plants but hyperaccumulator species may survive there. The critical toxicity of soils for plants has been defined using arbitrary thresholds, for example, 100–300 $\mu g\ g^{-1}$ for Zn and 6–8 $\mu g\ g^{-1}$ for Cd (Krämer, 2010). Bert et al. (2002) proposed that soils with more than 300 $\mu g\ g^{-1}$ Zn or 2 $\mu g\ g^{-1}$ Cd be classified as metalliferous (or metal-contaminated) soils according to the French agricultural recommendation. Hyperaccumulator plants transport heavy metals from roots to shoots and can cope with high concentration of these heavy metals in soils. Hyperaccumulation in plants has been defined as the presence of $>$3,000 $\mu g\ g^{-1}$ Zn and of $>$100 $\mu g\ g^{-1}$ Cd in leaves (Krämer, 2010). *A. halleri* grows in both contaminated soils near mines and non-contaminated soils, and plants from both types of soils can hyperaccumulate heavy metals (Bert et al., 2002; Stein et al., 2017). Because hyperaccumulation is constitutive in *A. halleri*, this species may have evolved this characteristic as a mechanism to extract high amounts of heavy metals from metal-deficient soils for chemical defense (Boyd, 2007). Experimentally, *A. halleri* plants treated with Cd or Zn were more resistant to specialist and generalist insect herbivores (Kazemi-Dinan et al., 2014, 2015). The other diploid parent, *A. lyrata*, has known adaptations to serpentine soils, i.e., high concentration of magnesium (Mg) and nickel (Ni) in some of its regions of distribution. This appears to be local adaptation (Turner et al., 2010; Arnold et al., 2016) rather than a constitutive trait. While it is shown that some genotypes of *A. kamchatica* can accumulate substantial amounts of heavy metals under experimental conditions (Paape et al., 2016), the combination of laboratory data with those obtained *in natura* is necessary to study ecologically relevant natural variation (Shimizu et al., 2011; Yamasaki et al., 2017). Specifically, very little is known about whether *A. kamchatica* hyperaccumulates in field conditions, whether its habitats encompass soils with heavy-metal by natural or artificial processes, or how much quantitative variation in hyperaccumulation exists in the species.

The genetic basis of Zn and Cd hyperaccumulation has been studied extensively in *A. halleri* using comparative transcriptomics (Filatov et al., 2006; Talke et al., 2006), using QTL mapping in *A. halleri* and *A. lyrata* crosses (Courbot et al., 2007; Willems et al., 2007; Frérot et al., 2010) and functional genetics (Hanikenne et al., 2008). In *A. kamchatica*, these genes are inherited as homeologs from both diploid progenitors. The relative expression levels of both homeologs may be similar to the parental species ("parental legacy"), resulting in an expression bias (Buggs et al., 2014; Yoo et al., 2014) of genes involved in hyperaccumulation that were inherited from *A. halleri*. A bias in important heavy metal transporters among widespread genotypes of *A. kamchatica* would suggest that parental legacy or constitutive expression has been maintained throughout the species distribution.

Here we focus on heavy metal hyperaccumulation in *A. kamchatica* as an ecologically relevant quantitative trait, and aimed to answer the following questions. (1) Do the habitats of *A. kamchatica* contain high concentration of metals (Zn, Cd, Mg, and Ni) and is there evidence of artificially or naturally generated metalliferous soils? (2) Does hyperaccumulation of Zn and Cd occur in natural populations of *A. kamchatica*? (3) How much quantitative variation in hyperaccumulation exists in *A. kamchatica*? (4) Is *A. halleri* a more efficient hyperaccumulator than *A. kamchatica* at all concentrations of Zn? (5) Do genes involved in heavy metal transport and detoxification show a bias in the expression ratios of homeologs among *A. kamchatica* genotypes that may be derived from *A. halleri*? By combining field data with phenotyping in experimental conditions and homeolog-specific gene expression estimates, we provide the first range-wide study of constitutive heavy metal hyperaccumulation in an allopolyploid species.

MATERIALS AND METHODS

Plant and Soil Material From Natural Populations

Arabidopsis kamchatica (Shimizu et al., 2005) is an allotetraploid species that is distributed in East Asia and North America. The diploid parental species *A. halleri* and *A. lyrata* each possess eight chromosomes ($2n = 2x = 16$) and the allopolyploid has $2n = 4x = 32$ chromosomes. Al-Shehbaz and O'Kane (2002)

reported that its habitats are gravelly slopes, forest, alpine regions, roadsides and flooded areas (under *Arabidopsis lyrata* subsp. *kamchatica* as a synonym). In contrast to *A. halleri*, it is not reported at sites with mining activities. Leaf tissues were collected from 40 *A. kamchatica* populations in Japan, Russia, and Alaska (United States), and soil samples from 38 corresponding locations were collected, from soil surface (0–1 cm depth) adjacent to a group of several *A. kamchatica* plants in each location, to quantify heavy metals in natural conditions (**Supplementary Table 1**). To identify sampling sites, we explored those natural populations by exploiting information on specimens that were available in museums and herbariums to cover a wide range of geographic regions, spanning altitudes from 30 to 2,949 m around three mountain chains in the Japanese Alps (Kenta et al., 2011). Additional samples were collected from Alaska, United States based on herbarium collection locations and previously reported sites (Shimizu-Inatsugi et al., 2009). For most populations, we collected leaf tissues from three or more plants. We reported values for leaf accumulation based on the mean and median of the replicates at each site. After visiting these natural populations, we noticed that some of the collection sites of *A. kamchatica* exhibited signs of human modification. Thus, populations were categorized into near construction or no construction as follows: (1) near construction populations are <10 m from buildings and fences on concrete base; or paved roads or riverside with concrete and/or asphalt, (2) no construction populations grew on native soil, free from concrete or asphalt or any obvious human modification. It is possible that populations not close to construction may have been cryptically altered by artificially occurring processes although it is not visible now. We also collected leaf tissues from *A. halleri* from two sites in Japan and two sites in Russia. Soil samples were collected from the two Japanese sites, which are known mine sites [Tada mine (TADA) and Omoidegawa (OMD), Japan; Briskine et al., 2017]. The Russian samples were collected from herbarium specimens; soils from these locations were not collected.

Hydroponic Plant Growth Experiments

We conducted hydroponic experiments to measure Zn accumulation in leaves and roots using natural genotypes of *A. kamchatica* and *A. halleri* from germplasm collected. We used three, nine or 19 *A. kamchatica* genotypes. Two genotypes of *A. halleri* genotypes are TADA collected from a known mine site (subsp. *gemmifera*, a parental taxon of *A. kamchatica*), and BOD collected from a non-mine site (subsp. *halleri*). We also included synthesized *A. kamchatica* generated in our lab using *A. halleri* and *A. lyrata* parental genotypes (Akama et al., 2014), which is expected to possess identical or very similar sequences as the parents. Seeds of *A. kamchatica* were germinated on phytoagar (0.8%) and a mixture of oligonutrients (25 μM H_3BO_3, 5 μM $MnCl_2$, 1 μM $ZnSO_4$, 0.5 μM $CuSO_4$, 50 μM KCl, and 0.1 μM Na_2MoO_4) and plated on a square (8 cm × 8 cm) petri dish until the seeds started to germinate (about 1 week). We used 1,000 μL pipet tip boxes (∼700 mL volume) as hydroponic chambers, so that seedlings could be grown in 0.5 mL thermo-PCR tubes using the 96-well insert (about 20 seedlings per box, for adequate spacing). The seedlings were then transplanted in 0.5 mL thermo-PCR tubes that were also filled with phytoagar solution and placed in the pipet tip boxes. The hydroponic solution was prepared according to Paape et al. (2016) and was composed of: 4 mM KNO_3, 1.2 mM $Ca(NO_3)_2$, 0.8 mM $MgSO_4$, 0.8 mM KH_2PO_4, 0.8 mM NH_4Cl, and 5 μM Fe(III)EDTA. A separate 1 L stock of oligoelements was prepared with the following elements (16.25 mL of oligonutrients was added to the final 5 L solution): 0.2 mM KCl, 0.12 mM H_3BO_3, 0.04 mM $MnSO_4$, 4 μM $CuSO_4$, 1 μM $ZnSO_4$, and 1 μM $(NH_4)_6Mo7O_2$. All Zn treatments were in the form of $ZnSO_4$ to ensure Zn is soluble in the hydroponic solution. Ten-liter batches were mixed in one container and then dispensed to individual hydroponic containers. The final pH was adjusted to 5.6–5.8.

To keep the moisture suitable for the growth of immature seedlings, a plastic bag was wrapped around each container to maintain a high level of humidity. The container was placed near natural light for 3–4 days. Plants that initially germinated but died in the immature stage were discarded. The final number of biological replicates for each accession varied from 5 to 13. Each hydroponic chamber contained a single plant genotype, to avoid root contamination between genotypes during sampling and harvesting. The boxes were then moved into a growth chamber (16 h light/8 h dark at 20°C) and the plastic bag was removed. The boxes containing the seedlings were placed on a 40 cm × 60 cm green tray (four boxes per tray) containing 1 cm of water and were covered with a plastic lid, to maintain a high level of humidity. The bottom of the plastic tube was cut with scissors after the roots grew to ∼0.5 cm, to allow the root to elongate into the box containing the hydroponic solution. A similar procedure was used for *A. halleri*, by placing freshly cut clones (i.e., ramets from a living plant) directly into the 0.5 mL tubes. Once the plants achieved the 3–4 leaf stage, the lid was removed to allow direct exposure to light. Zn supplements (Zn treatments added to the solution in the form of $ZnSO_4$) were added after about 4.5 weeks of plant growth (with slight variability among genotypes because of seed germination). Leaves and roots were harvested from plants after 5.5 weeks of growth, with exposure to Zn treatment during the final week of growth.

We performed three Zn treatment experiments. In the first experiment, a supplement of 500 μM $ZnSO_4$ was added to the hydroponic solution for 7 days. This experiment included 19 natural *A. kamchatica* genotypes and one synthetic polyploid that was generated from *A. halleri* subsp. *gemmifera* and *A. lyrata* subsp. *petraea* (**Supplementary Material**). We also used two naturally collected *A. halleri* genotypes that are maintained in our lab, TADA (*A. halleri* subsp. *gemmifera*, originally collected from the Tada mine, Japan) and BOD (*A. halleri* subsp. *halleri* collected from Boden, Switzerland), which can be clonally propagated for experiments. Because *A. kamchatica* is a self-fertilizing species, we can obtain offspring (seeds) from parents derived from single-seed descent. We grew 10–15 replicates per plant genotype, depending on germination.

The second experiment included nine *A. kamchatica* genotypes with a supplement of 1,000 μM Zn added to the hydroponic solution for 7 days. In the third experiment ("Zn gradient experiment"), Zn treatments were administered

at 10-fold increments of Zn concentration: 1 μM (control condition) and 10, 100, and 1,000 μM, to compare relative accumulation at different concentrations and between species and genotypes. We tested three *A. kamchatica* genotypes selected to represent different geographic regions: Alaska (PAK), Japan (MUR), and Sakhalin Island (SAK), and two *A. halleri* genotypes, one from the Tada mine area in Japan (TADA) and another from Boden, Switzerland (BOD). For *A. kamchatica* we grew 10–12 replicates each, for *A. halleri* we grew 6–8 replicates of each genotype. Plants were grown for ∼4.5 weeks prior to the treatments. Leaves and roots were harvested after 48 h exposure to each of the four treatments, to measure the short-term uptake of Zn.

Elemental Analysis in Plant Tissues and Soil Samples

Measurements of heavy metal concentrations were performed at the Institute of Terrestrial Ecosystems at ETH Zürich as described by Lahner et al. (2003) and Paape et al. (2016). For metal analysis in all three experiments, leaves and roots (approximately 2–5 mg dry weight) were harvested from plants after the Zn treatments. During harvesting, root tissues were washed in 150 mL of a cold solution of 5 mM $CaCl_2$ and 1 mM MES-KOH (pH 5.7) for 30 min, followed by a wash in 150 mL of cold water for 3 min. The tissues were then collected in paper envelopes and dried at 60°C for 2 days. Root tissues were rinsed again with 18 MΩ water and placed into Pyrex digestion tubes. Leaf and root tissues were dried at room temperature and then at 50°C for 24 h.

Plant tissues were weighed and placed into 50 mL tubes. Samples were placed into an oven at 92°C to dry for 20 h prior to ion measurement. After cooling, reference samples for each ion were weighed. Samples were digested in a microwave oven with 2 mL of concentrated nitric acid (HNO_3, ACS reagent; Sigma-Aldrich) and 30% hydrogen peroxide (Normapur; VWR Prolabo) and diluted to 10 mL with 18 MΩ water. Analytical blanks and standard reference material (WEPAL IPE 980) were digested together with plant samples. ICP-MS was used for the elemental analysis of samples and reference standards. To correct for instrumental drift, an internal standard containing yttrium and indium was added to the samples. All samples were normalized to calculate weights, as determined with a heuristic algorithm using the best-measured elements, the weights of the samples, and the elemental solution concentrations.

Soil samples were weighed and finely ground, then dried at 40°C. Soils were digested using the DigiPREP MS digestion system (SCP Science) and 2 M HNO_3 for 90 min at 120°C. The samples were then cooled and diluted with up to 50 mL of nanopure water. The digested soils were filtered using Whatman filter paper into 50 mL centrifuge tubes. Samples were diluted and measurements were performed using inductively coupled plasma optical emission spectrometry (ICP-OES) at ETH Zürich.

RNA Extraction and Pyrosequencing

Pyrosequencing of reverse-transcribed cDNA templates was used to measure the homeolog expression ratios of metal-ion transporter genes (see **Supplementary Material** for gene details). Pyrosequencing is a PCR-based method that can detect the relative abundance of homeolog-specific single-nucleotide polymorphisms (SNPs), which will vary according to homeolog expression levels (Akama et al., 2014). Leaf and root tissues from 15 *A. kamchatica* genotypes were harvested from three replicates grown in hydroponic solution at time zero (control) and at 48 h after the addition of 500 μM Zn (from the 500 μM Zn experiment described above). Leaf and root tissues were flash frozen in liquid nitrogen and stored at −80°C. RNA was extracted with TRIzol (Invitrogen) and purified using the RNeasy Mini Kit (Qiagen). RNA concentrations were measured using Nanodrop (Thermo Scientific). The RNA samples were reverse transcribed to cDNA using a High Capacity RNA-to-cDNA kit (Invitrogen).

Based on previous studies, we selected the orthologous *Arabidopsis thaliana* genes *HMA3* (AT4G30120), *HMA4* (AT2G19110), *MTP1* (AT2G46800), *MTP3* (AT3G58810), *NRAMP3* (AT2G23150), and *IRT3* (AT1G60960) (see **Supplementary Table 8** for functional details and citations). We used coding sequence alignments in FASTA format that were generated from resequencing data of *A. kamchatica* homeologs (Paape et al., 2018). The PyroMark Assay Design v2.0 software (Qiagen) at the Genomic Diversity Center (GDC), ETH Zürich, was used to design PCR primers, sequencing primers, and SNP assays for each gene. To design pyrosequencing assays, we aligned the coding sequences of homeologous gene copies for each of six known heavy metal transporter genes in *A. thaliana* (*HMA3*, *HMA4*, *IRT3*, *MTP1*, *MTP3*, and *NRAMP3*), to detect SNPs between copies derived from *A. halleri* (H-origin) and *A. lyrata* (L-origin). We searched for target SNPs between two conserved primers that contained 2–3 target SNPs in regions <200 bp using the PyroMark software. The amplified PCR fragments were sequenced using PyroMark Q96 ID. Amplification peaks were analyzed using the allele quantification (AQ) mode in the PyroMark software to determine the SNP amplification ratio. The ratios of the 2–3 target SNPs obtained for each gene fragment were averaged to estimate the H- and L-origin homeolog expression ratios. Standard deviations for each gene were estimated from the three biological replicates. For the H-origin homeologs of *HMA4* and *MTP1*, we assumed that the H-origin expression was the sum of the duplicated copies derived from *A. halleri* (Paape et al., 2016).

Statistical Analysis

For the leaf tissues and soils collected from natural populations, we calculated the mean, median, range, and standard deviations for each population, which typically consisted of three replicates. Correlations between leaf accumulation of Zn and Cd and soil concentrations of these heavy metals were assessed using Pearson's correlation coefficients. We used ANOVA and linear models (lm) to detect quantitative variation in Zn accumulation in leaf and root tissues among genotypes of *A. kamchatica* in the 500 μM Zn-treatment experiment. We estimated broad sense heritability (H^2) using the ANOVA table sum of square values to quantify between genotype variance relative to within genotype variance plus residual variance. We assessed whether the leaf and root Zn levels and the leaf/root ratio of Zn varied among

genotypes of *A. kamchatica*. We performed this analysis with and without the outlier genotype, MAG, to determine whether trait variation was affected. We used the linear model formula: Zn ∼ species + genotype (separately for leaf, and for root) as a statistical test to determine whether Zn accumulation was significantly different between *A. kamchatica* and *A. halleri*. For the Zn gradient experiment, we compared the means and standard deviations of three replicates of each genotype at each Zn treatment condition.

We used linear models to determine the effects of genotype, gene (function), treatment, and tissue on homeolog expression ratios (estimated using pyrosequencing). The expression ratios varied from 0 to 1 according to the relative expression of either homeolog. If there was equal expression of both homeologs, then H-origin = 0.5 and L-origin = 0.5. H-origin ratio is H-origin/(H-origin + L-origin). Therefore, we used the expression of the H-origin copy as the dependent variable. The following linear model formula, which included both tissue types, was used: H-origin ratio ∼ genotype + gene + treatment + tissue + genotype × gene (interaction term) + gene × treatment (interaction term) + gene × tissue (interaction term) + gene × treatment × tissue (interaction term). The significance of each explanatory variable was summarized by the sum of squares and *F*-statistics in an ANOVA table. All analyses were performed in R, version 3.4.

RESULTS

Zn Accumulation in Soils and Leaf Tissues in Wild Populations

We searched for populations of *A. kamchatica* in Japan and Alaska (United States) by using herbaria and database information (Shimizu-Inatsugi et al., 2009; Kenta et al., 2011). We sampled leaf tissues from 40 populations of *A. kamchatica* and we obtained soils from 38 of these localities (**Supplementary Tables 1–4**). The concentration of Zn in these soils ranged from 18.7 to 642.8 μg g^{-1} (average, 153.3 μg g^{-1}; **Figure 1** and **Supplementary Tables 1, 2**). Although we found no population close to active or historical mine sites with heavy metal contamination, consistent with previous reports (Al-Shehbaz and O'Kane, 2002), the Zn concentrations were >100 μg g^{-1} in soils from 18 sites and >300 μg g^{-1} in those from five sites. This indicated that several sites contained Zn above the critical toxicity of 100–300 μg g^{-1}, which would be toxic for most plants.

This surprisingly high Zn concentration in many populations suggest that these Zn may not be of natural origin but affected by human activities. We noticed that about half of these localities are near a human construction (**Supplementary Figure 1**). Then we classified the localities into near construction or no construction sites. The average Zn concentration in soils at near construction sites was significantly higher than that detected at no construction sites ($P = 0.00013$, **Supplementary Table 3** and **Figure 1**), suggesting artificial introduction of Zn into the habitats where human construction has occurred. Most of the sites with >300 μg g^{-1} Zn in soils were near construction (**Supplementary Table 1**). However, many sites with no construction also showed considerable levels of Zn (five sites with >100 μg g^{-1} Zn) and are likely to reflect the natural geology (Schlüchter et al., 1981), although it is possible that unobvious previous human activities may have affected, too. These data suggest that the habitat of *A. kamchatica* encompasses soils with high concentrations of Zn due to natural geology or human modification, as well as sites in either category that have low Zn.

The average accumulation of Zn in leaves was 1,416 μg g^{-1} dry weight among 126 plants from the 40 populations. We detected >1,000 μg g^{-1} Zn in 21 of the 40 populations and >3,000 μg g^{-1} Zn in four populations from both near construction and no construction sites (**Figure 1** and **Supplementary Table 1**). The highest leaf concentration of Zn was found in three plants at the Mt. Sirouma site (6,500–6,661 μg g^{-1}, **Supplementary Table 2**) in Japan (2,835 m above sea level), a site near construction with the highest level of Zn in the soil (**Figure 1**). We found that plants from near construction habitats had, on average, significantly higher quantities of Zn (mean Zn concentration, 1,808 μg g^{-1}) than did no construction sites (mean Zn concentration, 926 μg g^{-1}) (**Supplementary Table 3**, $P = 0.0005$, **Figure 1**). In near construction habitats, there was a significant positive correlation between the leaf and soil concentrations of Zn ($r = 0.67$, $P < 10^{-5}$), suggesting that the increased concentrations of Zn observed in the soils at these sites increased the availability of the metal for uptake by plants. In contrast, there was no correlation between the leaf and soil concentrations of Zn in no construction habitats ($r = -0.007$, $P = 0.95$; **Supplementary Figure 1**).

In the parental species *A. halleri*, collected from two sites in Russia and two sites in Japan, we detected >3,000 μg g^{-1} Zn in the leaves of all four populations, with the highest levels found in plants from the Tada mine site in Japan [average of six replicates = 16,068 μg g^{-1} (**Supplementary Table 1**)]. Soils were collected at the two mine sites in Japan [Tada and Omoidegawa (OMD)] (**Supplementary Table 2**), and the Zn concentration (1,317–2,490 μg g^{-1} Zn) was several times higher than at any other site containing *A. kamchatica*.

Cd and Serpentine Soils in the Habitats of *A. kamchatica*

The soils from all sites had <2 μg g^{-1} Cd, indicating the absence of contamination with this heavy metal in the habitats of *A. kamchatica*. The average accumulation of Cd in leaves among the populations of *A. kamchatica* was 1.8 μg g^{-1} (**Figure 2** and **Supplementary Table 3**) and the Cd concentration in leaves in the five populations with the highest Cd levels ranged from 3.6 to 8.9 μg g^{-1}. The maximum concentration of Cd in any single plant was 12.1 μg g^{-1} (**Supplementary Tables 2, 3**) recorded at the Mt. Siraiwa site, which also had the highest Cd levels in soils. The level of Cd accumulation in leaves was much lower than the threshold defined for Cd hyperaccumulation (>100 μg g^{-1}) (Krämer, 2010). Unlike Zn, there was no significant difference in Cd concentration in

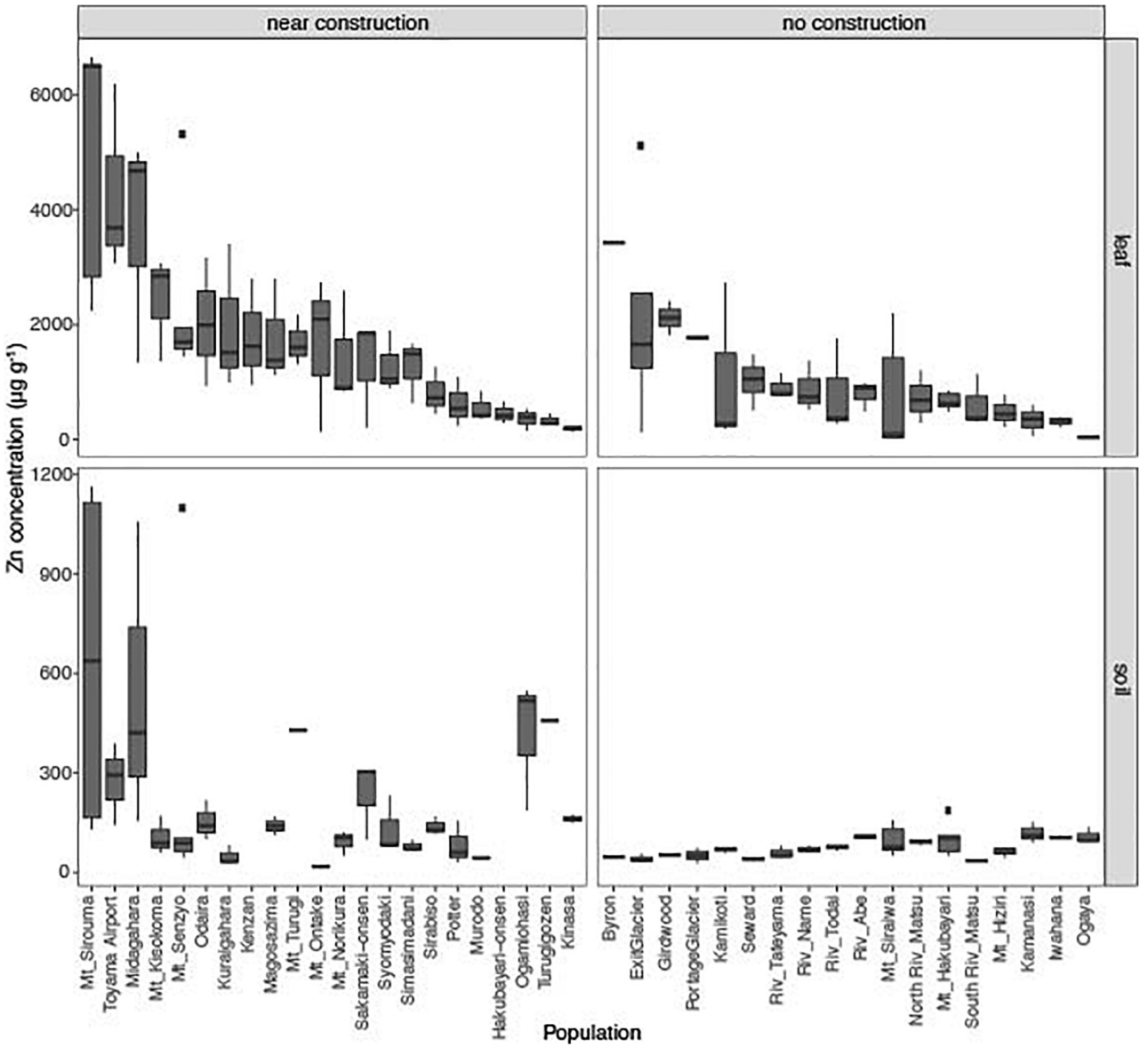


FIGURE 1 | Leaf concentrations of zinc (Zn) from 40 natural populations of *A. kamchatica* **(Upper)**. Soil concentrations of Zn **(Lower)** at 38 corresponding sites. Populations were grouped into near construction or no construction types. The populations are sorted from highest to lowest Zn concentration in the leaf samples. The *y*-axis is the Zn concentration in µg per gram of dry weight. Note different *y*-axis scales for leaf and soil samples. Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers. The correlations between leaf accumulation and soil concentrations: Pearson's $r = 0.68$ (near construction), Pearson's $r = 0.007$ (no construction) (see also **Supplementary Figure 1**). Refer to **Supplementary Table 1** for mean values, variance, number of replicates, and location information.

the soils or leaf tissues between the near construction or no construction habitats (**Supplementary Table 3**). Moreover, there was no correlation between the leaf and soil concentrations of Cd in the near construction habitats, but there was a positive correlation in the no construction habitats (**Supplementary Figure 1**). We also found only a weak correlation between Zn and Cd leaf accumulation ($r = 0.08$, $P = 0.39$), likely because of the much lower overall Cd concentration in soils and leaves compared with Zn. In summary, the leaf and soil concentrations of Cd were below the critical toxicity levels for plants and were negligible compared with Zn concentrations in the same populations.

Although the main objective of the soil and leaf sampling of *A. kamchatica* was to quantify the heavy metals Zn and Cd, we also found high levels of magnesium (Mg) and nickel (Ni) in the soils of two populations from Japan (Mt. Sirouma and Mt. Hakubayari) (**Supplementary Figures 2, 3** and **Supplementary Table 4**). These two sites had nearly an order of magnitude greater Mg and Ni levels than those observed for all other populations [the Mt. Sirouma site also contained the

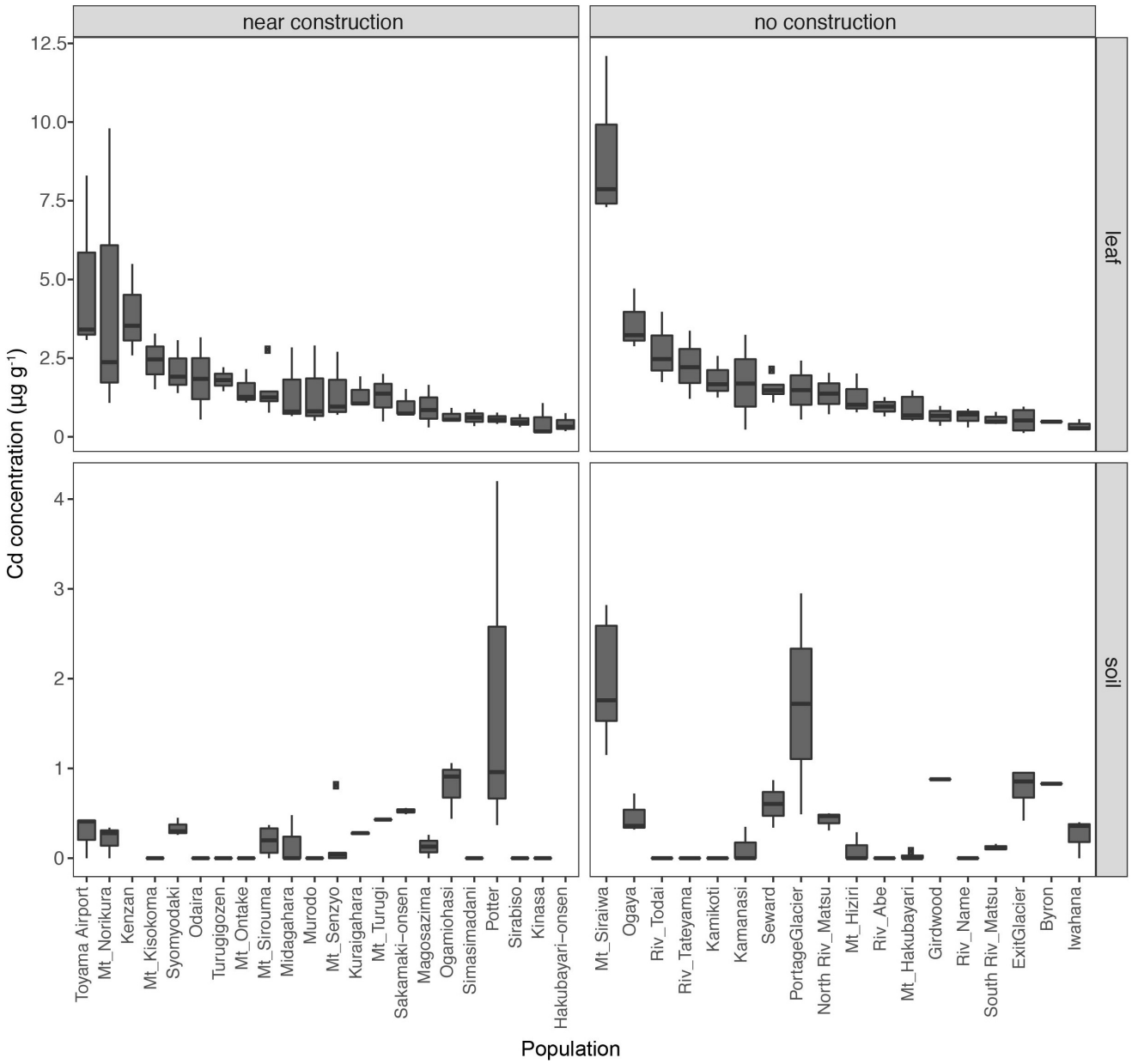


FIGURE 2 | Leaf concentrations of cadmium (Cd) in 40 natural populations of *A. kamchatica* **(Upper)**. Soil concentrations of Cd **(Lower)** at 38 corresponding sites. Populations were grouped into near construction or no construction types. The populations are sorted from highest to lowest Cd concentration in the leaf samples. The *y*-axis is the Cd concentration in μg per gram of dry weight. Note the different *y*-axis scales for leaf and soil samples. Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers. The correlations between leaf accumulation and soil concentrations: [near construction Pearson's $r = -0.08$, (no construction) Pearson's $r = 0.58$ (see **Supplementary Figure 1**)]. Refer to **Supplementary Table 1** for mean values, variance, number of replicates, and location information.

highest soil concentration of Zn among all of the *A. kamchatica* sites (643 μg g^{-1})]. High concentrations of Mg and Ni and low calcium-to-magnesium ratios (Ca:Mg) indicate serpentine soils (Kazakou et al., 2008). Consistent with this, the Ca:Mg in soils from Mt. Sirouma (Ca:Mg = $3.71e^{-05}$) and Mt. Hakubayari (Ca:Mg = $3.45e^{-05}$) were at least an order of magnitude lower than the average value among all other Japanese populations (on average $1.60e^{-03}$, sd. $2.60e^{-03}$). Moreover, the concentrations of Ni in the leaf tissues collected in the Mt. Sirouma, Mt. Hakubayari, and North and South River Matu sites were the highest among all Japanese populations (**Supplementary Figure 3**), and Mg concentration was highest among the leaf tissues from the Mt. Hakubayari and North and South River Matu populations (**Supplementary Figure 2**). Therefore, the high levels of Mg and Ni and the low Ca:Mg in soils and leaves at these sites indicate that *A. kamchatica* lives on serpentine soils;

in fact, the broader mountain range that includes Mt. Sirouma and Mt. Hakubayari has serpentine soils (Hatano and Matsuzawa, 2008), and the North and South River Matu originates in this mountain range where runoff can deposit Mg and Ni to lower elevations.

Variation in Zn Accumulation in Experimental Conditions

To examine the quantitative variation of Zn hyperaccumulation in *A. kamchatica* in a common environment, Zn accumulation in leaves and roots was quantified using hydroponic growth chambers. We used a treatment condition of 500 μM Zn for 7 days. Previously, we treated four genotypes of *A. kamchatica*, one *A. halleri* genotype and one *A. lyrata* genotype, and reported that the leaf Zn concentration of *A. kamchatica* was much higher than that of *A. lyrata* but about half of *A. halleri* (Paape et al., 2016). Here we used 20 *A. kamchatica* as well as 2 *A. halleri* genotypes. Among the 20 plant genotypes tested, the average Zn accumulation in leaves was 4,562 μg g^{-1} and the mean values ranged from 1,845 to 16,213 μg g^{-1} among replicates of all genotypes (**Figure 3** and **Supplementary Table 5A**). Linear models detected significant variation in Zn accumulation, even when the outlier genotypes were removed (**Supplementary Table 5B**). The broad-sense heritability (H^2) for leaf accumulation of Zn in *A. kamchatica* was estimated to be as high as 0.70. In addition, we treated nine of the 20 *A. kamchatica* genotypes with a higher concentration of Zn (1,000 μM) and found significant increases in Zn accumulation in the leaves in all but two genotypes

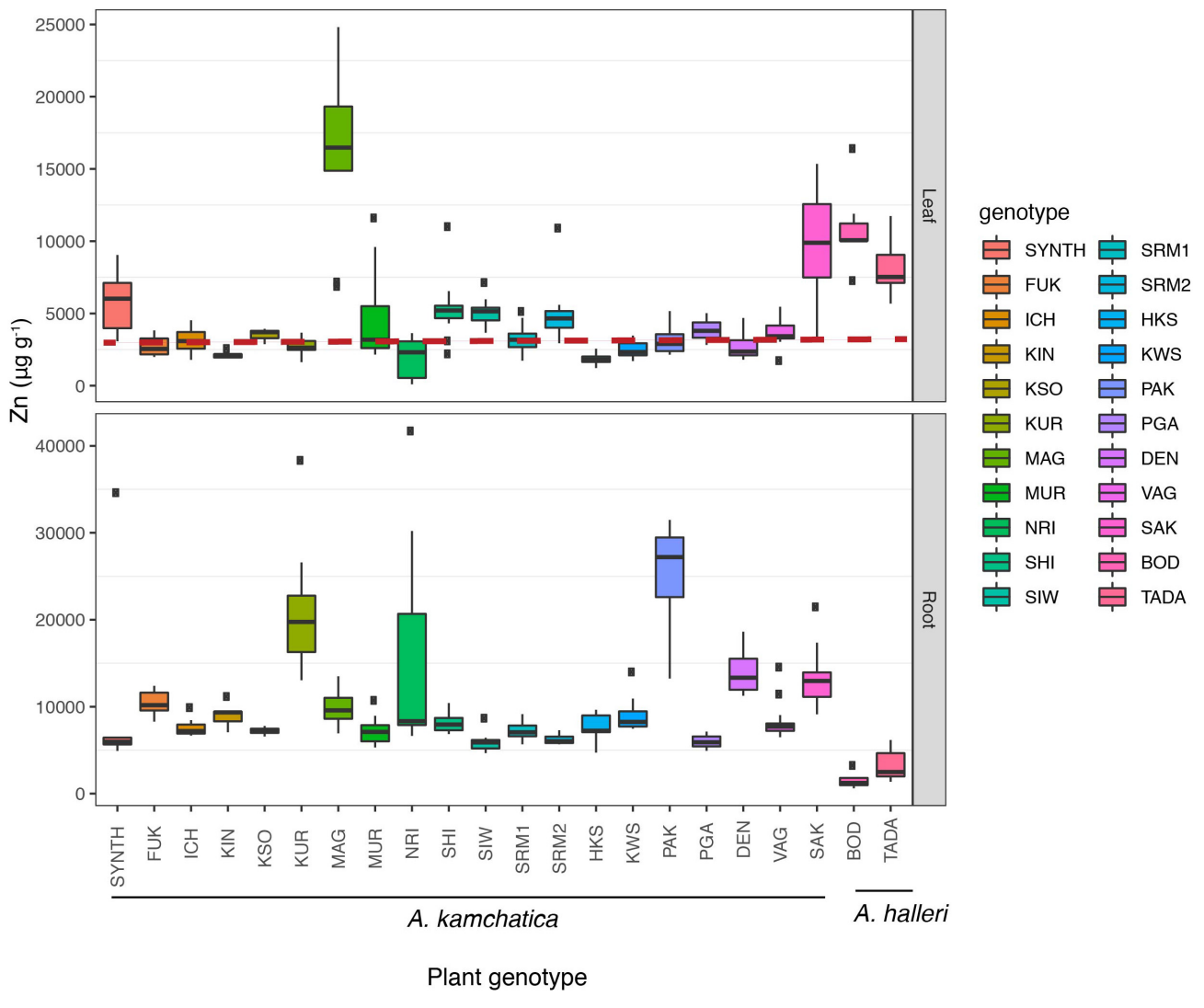


FIGURE 3 | Zinc accumulation in leaf tissues **(Upper)** and root tissues **(Lower)** in 20 *A. kamchatica* and two *A. halleri* genotypes after 1 week of 500 μM zinc treatment in hydroponic chambers. The dashed red line represents 3,000 μg. Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers. Boxplots are colored according to plant genotype. Refer to **Supplementary Table 5** for mean and median values, standard deviations, and number of replicates.

(**Supplementary Figure 4**). To check whether Zn accumulation among these samples was correlated with the soil concentrations of Zn in our field-collected samples, we compared 9 populations of *A. kamchatica* for which both soil Zn concentration and leaf accumulation in the growth chamber experiment were available. The variance of Zn concentration in leaf tissues that was explained by Zn concentration in soil was not significant ($P = 0.8$); this variance could be explained mostly by plant genotype ($P < 2e^{-16}$).

For comparison, we included two *A. halleri* genotypes in the 500 μM Zn experiment. The TADA and BOD genotypes accumulated 8,036 and 10,891 μg g^{-1} Zn in leaf tissues, respectively (**Figure 3** and **Supplementary Table 5**). While the average Zn accumulation in the leaves of *A. kamchatica* is about half of *A. halleri*, two natural genotypes (the MAG and SAK genotypes) accumulated Zn in leaves at similar or higher levels compared with *A. halleri*. A synthetic allopolyploid generated from *A. halleri* ssp. *gemmifera* and Siberian *A. lyrata* ssp. *petrea* (Akama et al., 2014) was included for comparison with natural polyploid genotypes and the TADA genotype. The synthetic polyploid showed a higher level of Zn accumulation in leaves than most natural genotypes (5,845 μg g^{-1}). However, this value was significantly lower than the *A. halleri* parental genotype (TADA) ($P = 0.04$, **Supplementary Table 6**). In contrast to natural *A. kamchatica* which experienced evolutionary changes after polyploid speciation such as mutation and gradual attenuation of expression over time due to environmental conditions, the synthetic allopolyploid provided direct experimental evidence that Zn hyperaccumulation from *A. halleri* can be inherited by *A. kamchatica*, but that the trait is reduced by the allopolyploidization.

The major difference in Zn accumulation between *A. kamchatica* and *A. halleri* was in the roots. The Zn concentrations in the roots of the 20 *A. kamchatica* genotypes were 5,980–25,650 μg g^{-1}, whereas Zn concentration in the roots of *A. halleri* was 3,190 μg g^{-1} for the TADA genotype and 1,525 μg g^{-1} for the BOD genotype; this was lower than all *A. kamchatica* genotypes (**Figure 3** and **Supplementary Table 5A**). The shoot/root ratio of Zn accumulation was >1 (i.e., higher Zn concentrations in leaves vs. roots) for both genotypes of *A. halleri*, while all but one of the *A. kamchatica* genotypes had a leaf-to-root ratio of Zn accumulation <1 [the MAG genotype had a leaf-to-root ratio >1 (**Supplementary Table 5A**)]. Pairwise comparisons of leaf accumulation showed that nearly all *A. kamchatica* genotypes, except MAG and SAK, differed significantly from both *A. halleri* genotypes (**Supplementary Table 6A**). Using a linear model with Zn accumulation as the dependent variable and species (*A. kamchatica* and *A. halleri*) as the explanatory variable (**Supplementary Table 6B**), we found that Zn accumulation in leaves was significantly greater in *A. halleri* ($p = 0.028$) and Zn accumulation in roots was significantly greater in *A. kamchatica* ($p = 0.00067$).

The long duration of the 500 μM Zn treatment (1 week) may have resulted in Zn transport approaching equilibrium levels in *A. kamchatica*. With sufficient time, the polyploid could potentially accumulate levels of Zn similar to those of *A. halleri*. To obtain a better picture of the relative efficiency of the physiological transport of Zn from roots to shoots by *A. kamchatica* and *A. halleri*, we measured Zn accumulation over a 10-fold gradient of concentrations within a short time period (48 h). We tested Zn accumulation in the leaves and roots of two *A. halleri* genotype representing a mine-site (TADA) and a non-mine site (BOD) and three *A. kamchatica* genotypes (MUR, PAK, and SAK). In general, the concentration of leaf Zn is lower in *A. kamchatica* than in *A. halleri*, and that of root Zn is higher, although the difference among genotypes were large. The two *A. halleri* genotypes show considerable differences in leaf and root accumulation of Zn, consistent with previous report of differences between mine and non-mine sites (Stein et al., 2017). At basal concentrations of Zn (1 μM), the BOD genotype accumulated significantly more than the TADA genotype (**Figure 4A**), with more similar levels of accumulation between the two genotypes at higher concentrations. In the roots, the BOD genotype accumulated significantly more Zn at the 10–1,000 μM treatments than TADA (**Figure 4B**), resulting in a lower shoot-to-root ratio at

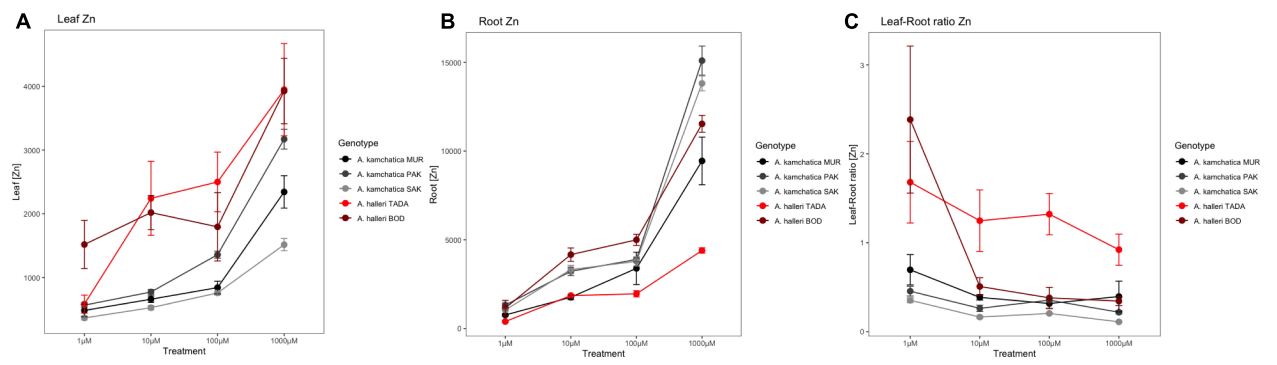


FIGURE 4 | Accumulation of Zn in the leaves **(A)** and roots **(B)** of *A. halleri* and *A. kamchatica* over a 10-fold gradient of Zn concentrations. Points and lines are color coded as follows: *A. halleri* (TADA, red) and three *A. kamchatica* genotypes (PAK, light gray; SAK, dark gray; and MUR, black) at four Zn treatment conditions (1, 10, 100, and 1,000 μM). The ratio of leaf-to-root Zn accumulation at each treatment condition is shown in **(C)**.

these treatments (**Figure 4C**). The shoot-to-root ratio of Zn concentration in each of the three treatment conditions was ≥1 for the TADA *A. halleri* genotype while the ratio is <1 for the BOD genotype (**Figure 4C**).

Among the three *A. kamchatica* genotypes, we found variation in Zn accumulation following the 10–1,000 μM treatments where the MUR genotype accumulated the highest amount of Zn and the SAK genotype had the lowest accumulation in leaves after the three treatments. Root tissues in these *A. kamchatica* genotypes accumulated Zn at levels between the two *A. halleri* genotypes, except at 1,000 μM where the PAK and SAK genotypes accumulated more Zn than both *A. halleri* genotypes. The high levels of Zn accumulation in the roots of *A. kamchatica* resulted in shoot-to-root ratios ≤0.5 at all Zn concentrations, which was significantly less than the TADA *A. halleri* genotype at each treatment level, but the shoot-to-root ratios are similar to the BOD *A. halleri* genotype (**Figure 4C**). These comparisons demonstrated that Zn accumulation is leaves of *A. kamchatica* is less than *A. halleri*, but that shoot-to-root ratios are within range of *A. halleri*.

Homeolog Expression Ratios of Candidate Genes for Heavy Metal Transport

We found an overall trend toward higher expression of *A. halleri*-derived (H-origin) homeologs compared with the *A. lyrata*-derived (L-origin) homeologs for all of the genes tested, although considerable variation was observed (**Figure 5**; see **Supplementary Table 8** for gene function information). Linear models showed that gene, plant genotype, and tissue contributed to a significant proportion of the variance in expression ratios ($P < 2.2e^{-16}$, $3.52e^{-13}$, and $1.02e^{-10}$, respectively; **Supplementary Table 9**). The genes *IRT3*, *MTP3*, and *NRAMP3* exhibited significant variation among the genotypes ($P = 7.15e^{--07}$, $3.89e^{-14}$, and $1.72\ e^{-11}$, respectively), while the *HMA3*, *HMA4*, and *MTP1* genes showed no significant among-genotype variation in expression ratios (P = 0.72, 0.72, and 0.21, respectively).

Zinc treatment had no significant effect on gene expression ratios compared with control conditions when both leaf and root tissues were included in the model (treatment P = 0.64) or when tissues were analyzed separately (leaf tissue only, P = 0.66; root tissue only, P = 0.12). These results reflect the constitutive expression of the H-origin metal transporters. Furthermore, we found no significant relationship between homeolog expression ratios and Zn hyperaccumulation levels in our experiments. The model coefficient for the interaction between gene and treatment was non-significant (gene × treatment interaction, P = 0.18) and the coefficients for each individual gene and treatment interaction were also non-significant (**Supplementary Table 9**). The only exception was the *MTP3* gene × treatment × tissue (root) interaction, which exhibited a significant coefficient (P = 0.0005). Whether this ratio change was caused by the up-regulation of the L-origin copy or the down-regulation of the H-origin copy cannot be determined by pyrosequencing, but the pattern is consistent with the up-regulation of the L-origin copy observed in a previous RNA-seq experiment (Paape et al., 2016).

DISCUSSION

Metalliferous and Non-metalliferous Soils and High Zn Accumulation in the Leaves of *Arabidopsis kamchatica*

In this first species-wide survey of *A. kamchatica*, ion measurements in leaf samples collected from plants growing in natural conditions showed that the species can accumulate large amounts of heavy metals. In particular, Zn accumulation was high in many populations, despite the generally low heavy metal concentrations in soils for the majority of sites. Substantial concentrations of Zn in leaves were found in more than half of the populations (>1,000 μg g^{-1}), and plants from four populations had >3,000 μg g^{-1} Zn, a threshold used to define hyperaccumulation in plants growing outside (Krämer, 2010). It is noteworthy that recent experiments have shown that *A. halleri* plants that accumulated up to 1,000 μg g^{-1} of Zn experienced a significant reduction in herbivory compared with plants with lower Zn accumulation (Kazemi-Dinan et al., 2014, 2015). Therefore, a similar level of Zn in leaves could also be sufficient for deterring herbivory in *A. kamchatica*.

Many of the habitats of *A. kamchatica* were metalliferous based on the criteria by Bert et al. (2002) but not due to mining, while the majority would be considered non-metalliferous. This was a major difference from *A. halleri* where many sites were highly contaminated because of mining activities (Pauwels et al., 2006; Hanikenne et al., 2013; Briskine et al., 2017), including two sites containing *A. halleri* used in this study (TADA and OMD). The influence of the *A. lyrata* genome may have reduced the ability of *A. kamchatica* to inhabit highly toxic mine sites. Nevertheless, we found that nearly half of the sites from which samples were collected were clearly modified by human activities such as construction, and the use of corrugated galvanized iron and gratings may have artificially increased Zn levels in the surrounding soils near mountain lodges and roads. Several of these sites contained high concentrations of Zn in the soils, and the Zn concentrations in the leaves near construction were significantly higher than those with no construction. This demonstrated that greater availability of Zn in the soils tends to result in plants that have higher levels of Zn in the leaves as we found in many of the modified sites. It is possible that inheritance of hyper-tolerance to heavy metals from *A. halleri* pre-adapted *A. kamchatica* to expand into human-modified sites that contained elevated Zn in soils.

Compared with Zn, the levels of Cd in soils or in field-collected *A. kamchatica* were unremarkable and below the toxic thresholds. There appeared to be no anthropogenic influence on Cd in the soils, as quantities at both human modified (near construction) and non-modified (no construction) sites differed only slightly and not significantly. Pollution from mining activities is the most common source of high amounts of Cd in soils and

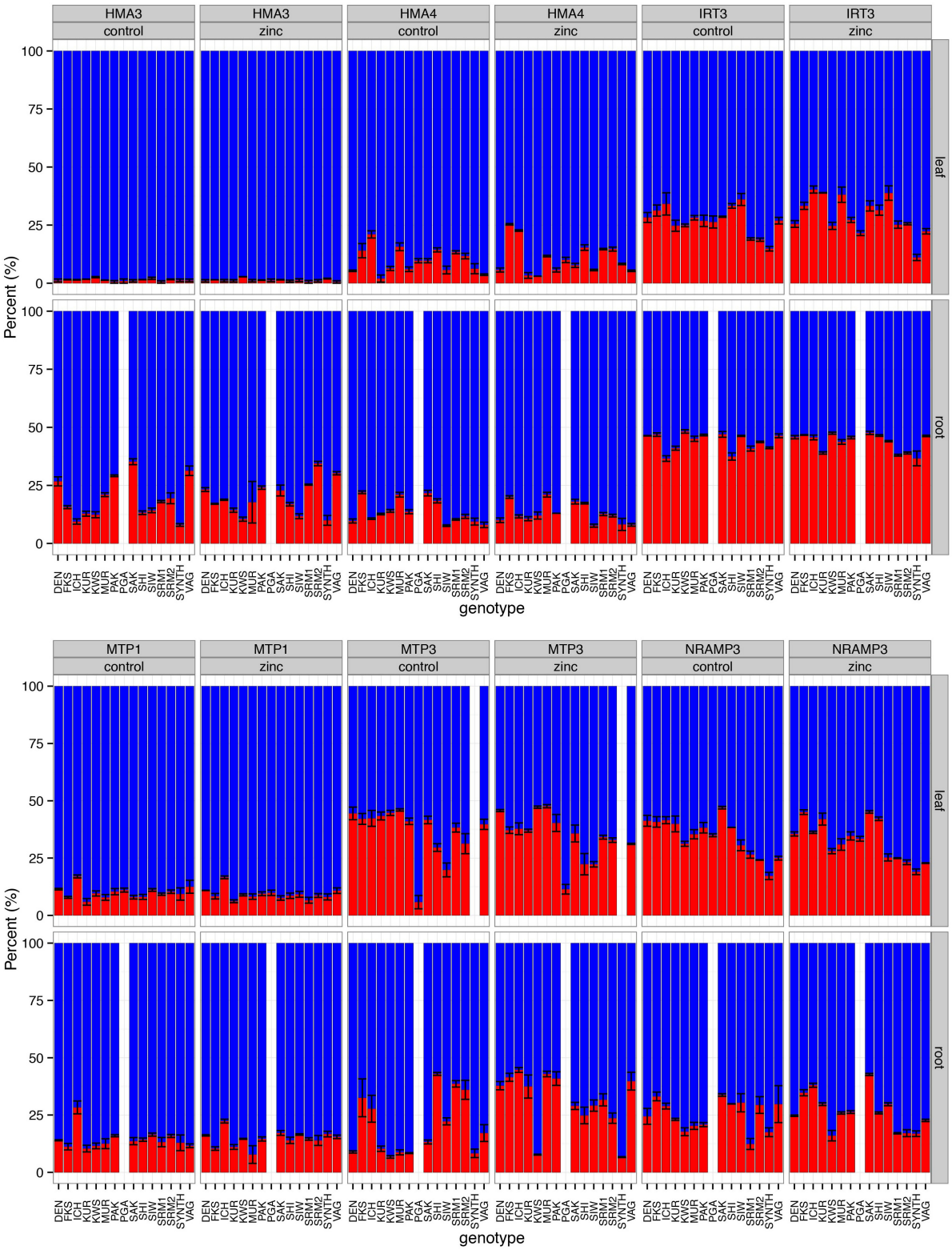


FIGURE 5 | The ratios of homeolog expression for six genes were estimated by pyrosequencing. Each bar on the horizontal axis represents 15 *A. kamchatica* genotypes (14 for root tissues) under control conditions. The vertical axis is the percentage of *A. halleri*-derived homeologs (blue) and *A. lyrata*-derived homeologs (red). For each gene, we estimated expression ratios under control (1 μM Zn) and zinc-treatment (500 μM Zn) conditions; plants were sampled 48 h after the administration of the treatment. Root data were not collected for the PGA genotype (data used to plot bar graphs can be found in **Supplementary Table 10**).

plants (Alloway and Steinnes, 1999), and elevated Cd from human construction is not a likely source of increased Cd. The correlation between Zn and Cd concentrations in leaf tissues of *A. kamchatica* plants was low, likely because of a lower exposure of plants to Cd, even among the populations that grew in soils with higher Zn concentrations.

Experimental Treatments Show That Zn Accumulation Is a Constitutive Trait in *A. kamchatica*

Because the range of concentrations of Zn in field-collected leaf samples varied by more than two orders of magnitude (from 30 to >6,000 $\mu g\ g^{-1}$), defining a species as a hyperaccumulator based on a single set threshold may be too simple for the characterization of intraspecific variation from samples collected in natural environments. Therefore, we used experimental treatments to quantify genotypic variation, so that the same exposure to Zn was provided to all plants. Our experiments detected an eightfold difference between genotypes for the lowest and highest Zn accumulation in leaf tissues, indicating a significant variation of Zn hyperaccumulation in *A. kamchatica*. Despite this variation under uniform treatments, the average Zn accumulation in leaves among the *A. kamchatica* genotypes was above 4,000 $\mu g\ g^{-1}$. We interpret this as a clear demonstration that *A. kamchatica* has constitutive hyperaccumulation ability as it is understood in the diploid parental species *A. halleri* (Hanikenne et al., 2008; Stein et al., 2017). Further support for constitutive Zn hyperaccumulation may be provided by the lack of a significant relationship between Zn concentrations in native soils and leaf accumulation levels in experimental conditions using *A. kamchatica* germplasm from the sites where soils were collected. Similarly, even with a large sampling of *A. halleri* genotypes from non-metalliferous sites, Stein et al. (2017) found no significant correlation between Zn concentrations in native soils and plants from the sites that were grown experimentally using Zn-amended soils (see Figure 3C in Stein et al., 2017).

We directly tested the inheritance of hyperaccumulation using a synthetic allopolyploid generated from *A. halleri* and *A. lyrata*. The synthetic *A. kamchatica* exhibited 73% of the accumulation of Zn in leaves measured in the parental *A. halleri* strain used in the same experiment (**Figure 3** and **Supplementary Table 5A**). This was considerably higher than that observed for most natural genotypes, which have on average about 50% of the accumulation of Zn in leaves measured in *A. halleri*. This clearly demonstrated that hyperaccumulation can be retained after hybridization between divergent parental species, although the *A. lyrata* genome had a weakening effect (by 27%) on this trait when Zn accumulation in *A. kamchatica* is compared with *A. halleri*.

Testing Zn accumulation in leaf and root tissues over a gradient of Zn treatments allowed us to further examine the physiological capacity of Zn accumulation in *A. kamchatica* compared with the parent species *A. halleri*. This experiment demonstrated that the shoot-to-root ratio in three polyploid genotypes never exceeded 0.5 at any treatment, while the shoot-to-root ratio in *A. halleri* for the TADA genotype that originated in a toxic mine site in Japan was ≥1 at all treatments. By contrast, the BOD genotype showed a strong hyperaccumulation response in leaves at the 1 μM Zn treatment but had similarly low accumulation in roots compared to all other genotypes including the TADA *A. halleri* genotype, but had a more similar shoot-to-root ratio to *A. kamchatica* (<1) at the 10–1000 μM treatments. A similar experiment was conducted comparing *A. halleri* and *A. thaliana* (a non-hyperaccumulator) (Talke et al., 2006), where the shoot-to-root ratio in *A. halleri* was similar to our results, but the shoot-to-root ratio (∼0.05) in *A. thaliana* was an order of magnitude lower compared with the shoot-to-root ratio we measured in *A. kamchatica* in this study.

We propose two explanations for the reduction in heavy metal hyperaccumulation in *A. kamchatica* compared with the diploid hyperaccumulator parent *A. halleri*. First, allopolyploidization is expected to reduce or attenuate the net expression levels of metal transporters compared with the diploid parents. Because allopolyploidization results in a state of fixed heterozygosity of functionally duplicated gene copies (homeologs), expression of homeologs may be reduced compared with the diploid parents. Previously, RNA-seq analysis showed a reduction of ∼50% in the expression of several genes encoding heavy metal transporters was detected in the *A. halleri*-derived homeologs of *A. kamchatica* compared with the orthologous genes in *A. halleri* (Paape et al., 2016). This reduction in the expression levels of metal transporter genes was consistent with the ∼50% lower levels of Zn accumulation in the leaves of natural accessions of *A. kamchatica* compared with *A. halleri*. Fixed heterozygosity can result in a trait that resembles a balanced polymorphism with a semi-intermediate phenotype. It is important to note that although the hyperaccumulation trait in *A. kamchatica* was reduced by ∼50% compared with *A. halleri*, it was an order of magnitude greater than that of the non-hyperaccumulating parent *A. lyrata* (Paape et al., 2016). Second, inhibiting genetic factors derived from the *A. lyrata* parental genome likely contribute to the reduced phenotype in the polyploid compared with *A. halleri*. It is expected that these genetic factors prevent the transport of toxic heavy metals to leaf tissues, which limits their toxicity in leaves. This is an example of genomic antagonism resulting from the divergent parental genomes.

Expression Bias in *A. halleri*-Derived Homeologs

Because Zn hyperaccumulation was inherited from *A. halleri*, we expected that the homeolog expression ratios would show a pattern consistent with a parental legacy effect of gene expression (Buggs et al., 2014). Pyrosequencing revealed that the expression ratios of homeologs for six genes with roles in heavy metal hyperaccumulation or metal tolerance exhibited higher expression of the H-origin copy. This suggests that homeolog-specific expression is maintained by *cis*-regulatory differences (Shi et al., 2012; Yoo et al., 2014). The two main loci involved in heavy metal hyperaccumulation in *A. halleri* are the heavy metal ATPase 4 (*HMA4*) gene, which encodes an ATPase transporter protein, and the metal tolerance protein 1 (*MTP1*; also called *ATCDF1* or *ZAT1*) gene, which encodes a cation diffusion

facilitator (CDF) protein (Frérot et al., 2010). In *A. halleri*, *HMA4* is tandemly triplicated and *MTP1* has at least three copies (Hanikenne et al., 2008; Shahzad et al., 2010; Briskine et al., 2017). A single copy of these two genes is present in *A. lyrata* (Briskine et al., 2017; Paape et al., 2018). Both genes have significantly higher expression than their non-hyperaccumulating congeners because of *cis*-regulation and gene duplications (Hanikenne et al., 2008; Shahzad et al., 2010). The *HMA3* gene (which encodes an ATPase in the same class as *HMA4*) also has a putative role in the vacuolar sequestration of Zn (Morel et al., 2008), similar to *MTP1*, but is only found in a single copy in both *A. halleri* and *A. lyrata* (Paape et al., 2018). These features are consistent with constitutive gene expression inherited from the hyperaccumulating diploid parent, *A. halleri*, which would be essential for retaining the hyperaccumulation phenotype in the species-wide collection of *A. kamchatica* examined in this study.

The ZIP-transporter *IRT3* and natural-resistance-associated macrophage protein 3 (*NRAMP3*) genes encode iron (Fe) transporters that are coregulated by Fe and Zn and show significant expression differences between the hyperaccumulator *A. halleri* and the non-accumulators *A. thaliana* (Talke et al., 2006; Lin et al., 2009; Shanmugam et al., 2011) and *A. lyrata* (Filatov et al., 2006; Paape et al., 2016). The *IRT3* and *NRAMP3* homeologs also exhibited an H-origin expression bias, which was consistent with previous differential expression studies that compared *A. halleri* with *A. lyrata* or *A. thaliana* in Zn-treatment studies (Filatov et al., 2006; Talke et al., 2006); however, their direct role in Zn hyperaccumulation is less clear (Thomine et al., 2003). Moreover, both *IRT3* and *NRAMP3* showed a much larger variation in expression ratio compared with *HMA4* and *MTP1*, which may reflect a greater constraint on the constitutive expression of the latter two genes.

Most importantly, Zn treatment had no significant effect on the expression ratio of five of these six genes, demonstrating the *A. halleri*-derived constitutive expression of genes with known or putative roles in Zn hyperaccumulation. The gene *MTP3* was an exception, as it exhibited a significant change in homeolog ratios in the root tissues of many genotypes after Zn treatment. This expression-ratio change was most likely the result of the previously demonstrated upregulation of the L-origin homeolog in roots of one of the genotypes used in the current study (Paape et al., 2016). *MTP3* prevents heavy metal transport to the shoots by sequestering Zn in the vacuoles in the roots of *A. thaliana* (Arrivault et al., 2006). We assume that this gene plays a similar role in *A. lyrata*, and is therefore a potential *A. lyrata*-derived inhibiting factor that would contribute to the reduced leaf hyperaccumulation observed in *A. kamchatica*.

Evolutionary Scenario of an Allopolyploid Habitat Expansion

Adaptability and range expansion in polyploids have been debated for several decades (Stebbins, 1971; Soltis et al., 2014; Van de Peer et al., 2017). However, empirical examples of genetically tractable quantitative traits that have ecological relevance are lacking (Godfree et al., 2017). We suggest that the inheritance of hyperaccumulation from *A. halleri* conferred advantages instantaneously following polyploid speciation, which was estimated to have occurred ∼100,000 years ago (Paape et al., 2018), according to the following scenario. First, *A. kamchatica* became tolerant to soils with toxic levels of heavy metals that were present in the areas of growth of natural populations because of geological processes. Subsequently, during the past few thousand years, soils became contaminated by human activities, and the tolerance functioned as a pre-adaptation to modified environments (as proposed for *A. halleri*; Meyer et al., 2016). In contrast to *A. halleri*, and consistent with its attenuated Zn hyperaccumulation, *A. kamchatica* was not found in extremely contaminated sites, such as mines. We found a distinct, intermediate habitat and species distribution of *A. kamchatica* such as soils near mountain lodges and roads. This supports the importance of a fine-scale environment for the habitat differentiation of polyploid species (Akiyama et al., 2019) in addition to climatic gradients at a large geographic scale (Hoffmann, 2005). Furthermore, Zn concentration in the leaves of the majority of the natural *A. kamchatica* populations was $>1{,}000\ \mu g\ g^{-1}$, which was an effective level for insect defense in *A. halleri* (Kazemi-Dinan et al., 2014). In addition, we found *A. kamchatica* living on serpentine soils, which has also been reported for the other diploid parent, *A. lyrata* (Turner et al., 2010; Arnold et al., 2016). We hypothesize that *A. kamchatica* expanded its habitats by combining the heavy metal and serpentine tolerances from *A. halleri* and *A. lyrata*, respectively. Our study represents a promising example of the contribution of the inheritance of genetic toolkits for soil adaptation to the habitat expansion of an allopolyploid species. New genomic and transcriptomic capabilities in *A. kamchatica* combined with functional genetics (Yew et al., 2018) and self-compatibility (Tsuchimatsu et al., 2012) provide a unique opportunity to study the genetics of the edaphic and climatic adaptation of a polyploid species (Shimizu et al., 2011).

DATA AVAILABILITY STATEMENT

All datasets used in this study are included in this study are included in the accompanying **Supplementary Files.**

AUTHOR CONTRIBUTIONS

TP and KS conceived the study. TP and TC designed and performed the experiments and prepared samples for analyses. TP designed the pyrosequencing assays and TC performed the pyrosequencing experiments. TP, YO, AH, TK, and KS collected field samples. TP and TC constructed the datasets. TP and RA performed the statistical analyses. TP drafted the manuscript. TP, RA, and KS revised the final draft. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the European Union's Seventh Framework Programme for research Swiss Plant Fellows to TP,

the University Research Priority Program of Evolution in Action of the University of Zurich to TP and KS, a JST CREST grant (number JPMJCR16O3) to KS and TK, a Swiss National Science Foundation grant 31003A_182318 to KS, MEXT KAKENHI grant numbers 16H06469 and 18H04785 to KS, the Special Coordination Funds for Promoting Science and Technology from the MEXT, Japan to TK, an Inamori Foundation research grant to TK, and a JSPS Grant-in-Aid for Scientific Research (2277023) to TK.

ACKNOWLEDGMENTS

We thank Rainer Schulin and Bjoern Studer at the Institute of Terrestrial Ecosystems at ETH Zürich for the metal ion analysis facilities, technical support, and useful discussion, and Keitaro Tanoi and Rie Shimizu-Inatsugi for discussion, Diana Wolf and Naoki Takebayashi for discussion about Alaskan populations; and the Genetic Diversity Center (GDC) for the software and equipment used to perform pyrosequencing. Some of the field leaf and soil samples were provided by Hinako Kanai. An earlier version of this manuscript was released as a pre-print at https://www.biorxiv.org/, (Paape et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.565854/full#supplementary-material

REFERENCES

Adams, K. L. (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. *J. Hered.* 98, 136–141. doi: 10.1093/jhered/esl061

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

Akiyama, R., Sun, J., Hatakeyama, M., Lischer, H. E. L., Briskine, R. V., Hay, A., et al. (2019). Fine-scale ecological and transcriptomic data reveal niche differentiation of an allopolyploid from diploid parents in *Cardamine*. *bioRxiv[Preprint]*. doi: 10.1101/600783

Alloway, B. J., and Steinnes, E. (1999). "Anthropogenic Additions of Cadmium to Soils," in *Cadmium in Soils and Plants*, eds M. J. McLaughlin and B. R. Singh (Dordrecht: Springer), 97–123. doi: 10.1007/978-94-011-4473-5_5

Al-Shehbaz, I. A., and O'Kane, S. L. Jr (2002). Taxonomy and phylogeny of Arabidopsis (Brassicaceae). *Am. Soc. Plant Biol.* 1:e0001. doi: 10.1199/tab.0001

Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., et al. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 8320–8325. doi: 10.1073/pnas.1600405113

Arrivault, S., Senger, T., and Krämer, U. (2006). The Arabidopsis metal tolerance protein AtMTP3 maintains metal homeostasis by mediating Zn exclusion from the shoot under Fe deficiency and Zn oversupply. *Plant J.* 46, 861–879. doi: 10.1111/j.1365-313x.2006.02746.x

Bert, V., Bonnin, I., Saumitou-Laprade, P., Laguerie, P., and Petit, D. (2002). Do *Arabidopsis halleri* from nonmetallicolous populations accumulate zinc and cadmium more effectively than those from metallicolous populations? *New Phytol.* 155, 47–57. doi: 10.1046/j.1469-8137.2002.00432.x

Bert, V., Macnair, M. R., De Laguerie, P., Saumitou-Laprade, P., and Petit, D. (2000). Zinc tolerance and accumulation in metallicolous and nonmetallicolous populations of *Arabidopsis halleri* (Brassicaceae): zinc tolerance and accumulation. *New Phytol.* 146, 225–233. doi: 10.1046/j.1469-8137.2000.00634.x

Blaine Marchant, D., Soltis, D. E., and Soltis, P. S. (2016). Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytol.* 212, 708–718. doi: 10.1111/nph.14069

Boyd, R. S. (2007). The defense hypothesis of elemental hyperaccumulation: status, challenges and new directions. *Plant Soil* 293, 153–176. doi: 10.1007/s11104-007-9240-6

Briskine, R. V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., et al. (2017). Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol. Ecol. Resour* 17, 1025–1036. doi: 10.1111/1755-0998.12604

Buggs, R. J. A., Wendel, J. F., Doyle, J. J., Soltis, D. E., Soltis, P. S., and Coate, J. E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130354–20130354. doi: 10.1098/rstb.2013.0354

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Courbot, M., Willems, G., Motte, P., Arvidsson, S., Roosens, N., Saumitou-Laprade, P., et al. (2007). A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with hma4, a gene encoding a heavy metal ATPase. *Plant Physiol.* 144, 1052–1065. doi: 10.1104/pp.106.095133

Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Filatov, V., Dowdle, J., Smirnoff, N., Ford-Lloyd, B., Newbury, H. J., and Macnair, M. R. (2006). Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation: gene expression in segregating families. *Mol. Ecol.* 15, 3045–3059. doi: 10.1111/j.1365-294x.2006.02981.x

Frérot, H., Faucon, M.-P., Willems, G., Godé, C., Courseaux, A., Darracq, A., et al. (2010). Genetic architecture of zinc hyperaccumulation in *Arabidopsis halleri*: the essential role of QTL × environment interactions. *New Phytol.* 187, 355–367. doi: 10.1111/j.1469-8137.2010.03295.x

Godfree, R. C., Marshall, D. J., Young, A. G., Miller, C. H., and Mathews, S. (2017). Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. *R. Soc. Open Sci.* 4:170934. doi: 10.1098/rsos.170934

Hanikenne, M., Kroymann, J., Trampczynska, A., Bernal, M., Motte, P., Clemens, S., et al. (2013). Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genet.* 9:e1003707. doi: 10.1371/journal.pgen.1003707

Hanikenne, M., Talke, I. N., Haydon, M. J., Lanz, C., Nolte, A., Motte, P., et al. (2008). Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature* 453, 391–395. doi: 10.1038/nature06877

Hatano, H., and Matsuzawa, T. (2008). Serpentine soil environment and distribution of alpine plants in the shirouma mountain range. *Jap. J. Ecol.* 58, 199–204. doi: 10.1142/9789813274211_0008

Hoffmann, M. H. (2005). Evolution of the realized climatic niche in the genus: Arabidopsis (Brassicaceae). *Evolution* 59, 1425–1436. doi: 10.1111/j.0014-3820.2005.tb01793.x

Huynh, S., Broennimann, O., Guisan, A., Felber, F., and Parisod, C. (2020). Eco-genetic additivity of diploids in allopolyploid wild wheats. *Ecol. Lett.* 23, 663–673. doi: 10.1111/ele.13466

Kazakou, E., Dimitrakopoulos, P. G., Baker, A. J. M., Reeves, R. D., and Troumbis, A. Y. (2008). Hypotheses, mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. *Biol. Rev.* 83, 495–508.

Kazemi-Dinan, A., Barwinski, A., Stein, R. J., Krämer, U., and Müller, C. (2015). Metal hyperaccumulation in Brassicaceae mediates defense against herbivores in the field and improves growth. *Entomol. Exp. Appl.* 157, 3–10. doi: 10.1111/eea.12333

Kazemi-Dinan, A., Thomaschky, S., Stein, R. J., Krämer, U., and Müller, C. (2014). Zinc and cadmium hyperaccumulation act as deterrents towards specialist herbivores and impede the performance of a generalist herbivore. *New Phytol.* 202, 628–639. doi: 10.1111/nph.12663

Kenta, T., Yamada, A., and Onda, Y. (2011). Clinal variation in flowering time and vernalisation requirement across a 3000-M altitudinal range in perennial *Arabidopsis kamchatica* ssp.Kamchatica and annual lowland *Subspecies kawasakiana*. *J. Ecosyst. Ecography* s6:001. doi: 10.4172/2157-7625.S6-001

Krämer, U. (2010). Metal hyperaccumulation in plants. *Annu. Rev. Plant Biol.* 61, 517–534. doi: 10.1146/annurev-arplant-042809-112156

Lahner, B., Gong, J., Mahmoudian, M., Ellen, L. S., Khush, B. A., David, E. S., et al. (2003). Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana. *Nat. Biotechnol.* 21, 1215–1221. doi: 10.1038/nbt865

Levin, D. A. (2002). *The Role of Chromosomal Change in Plant Evolution*, Oxford. New York, NY: Oxford University Press.

Lin, Y.-F., Liang, H.-M., Yang, S.-Y., Boch, A., Clemens, S., Chen, C.-C., et al. (2009). Arabidopsis IRT3 is a zinc-regulated and plasma membrane localized zinc/iron transporter. *New Phytol.* 182, 392–404. doi: 10.1111/j.1469-8137.2009.02766.x

Meyer, C.-L., Pauwels, M., Briset, L., Godé, C., Salis, P., Bourceaux, A., et al. (2016). Potential preadaptation to anthropogenic pollution: evidence from a common quantitative trait locus for zinc and cadmium tolerance in metallicolous and nonmetallicolous accessions of *Arabidopsis halleri*. *New Phytol.* 212, 934–943. doi: 10.1111/nph.14093

Morel, M., Crouzet, J., Gravot, A., Auroy, P., Leonhardt, N., Vavasseur, A., et al. (2008). AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis. *Plant Physiol.* 149, 894–904. doi: 10.1104/pp.108.130294

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer Berlin Heidelberg.

Paape, T., Akiyama, R., Cereghetti, T., Onda, Y., Hirao, A., Kenta, T., et al. (2019). Experimental and field data support habitat expansion of the allopolyploid *Arabidopsis kamchatica* owing to parental legacy of heavy metal hyperaccumulation. *bioRxiv[Preprint]*. doi: 10.1101/810853

Paape, T., Briskine, R. V., Halstead-Nussloch, G., Rie, S. I., Kenta, T., Jun, S., et al. (2018). Patterns of polymorphism and selection in the subgenomes of the allopolyploid *Arabidopsis kamchatica*. *Nat. Commun.* 9:3909.

Paape, T., Hatakeyama, M., Shimizu-Inatsugi, R., Cereghetti, T., Onda, Y., Kenta, T., et al. (2016). Conserved but attenuated parental gene expression in allopolyploids: constitutive zinc hyperaccumulation in the allotetraploid *Arabidopsis kamchatica*. *Mol. Biol. Evol.* 33, 2781–2800. doi: 10.1093/molbev/msw141

Pauwels, M., FréRot, H., Bonnin, I., and Saumitou-Laprade, P. (2006). A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *J. Evol. Biol.* 19, 1838–1850. doi: 10.1111/j.1420-9101.2006.01178.x

Ramsey, J., and Ramsey, T. S. (2014). Ecological studies of polyploidy in the 100 years following its discovery. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130352–20130352. doi: 10.1098/rstb.2013.0352

Schlüchter, C., Heuberger, H., and Horie, S. (1981). New evidence for multiglaciation in the high mountains of Japan. I. new observations in Hakuba (Shirouma)-dake. *Proc. Jpn. Acad. Ser B Phys. Biol. Sci.* 57, 296–299. doi: 10.2183/pjab.57.296

Shahzad, Z., Gosti, F., Frérot, H., Lacombe, E., Roosens, N., Saumitou-Laprade, P., et al. (2010). The five AhMTP1 zinc transporters undergo different evolutionary fates towards adaptive evolution to zinc tolerance in *Arabidopsis halleri* R. Mauricio, ed. *PLoS Genet.* 6:e1000911. doi: 10.1371/journal.pgen.1000911

Shanmugam, V., Lo, J.-C., Wu, C.-L., Wang, S.-L., Lai, C.-C., Connolly, E. L., et al. (2011). Differential expression and regulation of iron-regulated metal transporters in Arabidopsis halleri and Arabidopsis thaliana - the role in zinc tolerance. *New Phytol.* 190, 125–137. doi: 10.1111/j.1469-8137.2010.03606.x

Shi, X., Ng, D. W.-K., Zhang, C., Comai, L., Ye, W., and Jeffrey Chen, Z. (2012). Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nat. Commun.* 3:950.

Shimizu, K. K., Fuji, S., Marhold, K., Watanabe, K., and Kudoh, H. (2005). Arabidopsis kamchatica (Fisch. *ex DC.*). K. Shimizu Kudoh and A. kamchatica subsp. kawasakiana (Makino). K. Shimizu & Kudoh, New Combinations. *Acta Phytotaxon. Geobot.* 56, 165–174.

Shimizu, K. K., Kudoh, H., and Kobayashi, M. J. (2011). Plant sexual reproduction during climate change: gene function in natura studied by ecological and evolutionary systems biology. *Ann. Bot.* 108, 777–787. doi: 10.1093/aob/mcr180

Shimizu-Inatsugi, R., Lihová, J., Iwanaga, H., Kudoh, H., Marhold, K., Savolainen, O., et al. (2009). The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Mol. Ecol.* 18, 4024–4048.

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., and Shimizu, K. K. (2016). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* 26, 193–207. doi: 10.1111/mec.13738

Soltis, D. E., Visger, C. J., and Soltis, P. S. (2014). The polyploidy revolution then.and now: stebbins revisited. *Am. J. Bot.* 101, 1057–1078. doi: 10.3732/ajb.1400178

Stebbins, G. (1971). *Chromosomal Evolution in Higher Plants*. London: Edward Arnold LTD, 87–89.

Stein, R. J., Höreth, S., de Melo, J. R. F., Syllwasschy, L., Lee, G., Garbin, M. L., et al. (2017). Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically structured in the emerging model *Arabidopsis halleri*. *New Phytol.* 213, 1274–1286. doi: 10.1111/nph.14219

Sun, J., Shimizu-Inatsugi, R., Hofhuis, H., Shimizu, K., Hay, A., Shimizu, K. K., et al. (2020). A recently formed triploid *Cardamine insueta* inherits leaf vivipary and submergence tolerance traits of parents. *Front. Genet.* doi: 10.3389/fgene.2020.567262

Talke, I. N., Hanikenne, M., and Kramer, U. (2006). Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiol.* 142, 148–167. doi: 10.1104/pp.105.076232

Thomine, S., Lelièvre, F., Debarbieux, E., Schroeder, J. I., and Barbier-Brygoo, H. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant J.* 34, 685–695. doi: 10.1046/j.1365-313x.2003.01760.x

Tsuchimatsu, T., Kaiser, P., Yew, C.-L., Bachelier, J. B., and Shimizu, K. K. (2012). Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. *PLoS Genet* 8:e1002838. doi: 10.1371/journal.pgen.1002838

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., and Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42, 260–263. doi: 10.1038/ng.515

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

Willems, G., Drager, D. B., Courbot, M., Gode, C., Verbruggen, N., and Saumitou-Laprade, P. (2007). The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri (Brassicaceae*): an analysis of quantitative trait loci. *Genetics* 176, 659–674. doi: 10.1534/genetics.106.064485

Yamasaki, E., Altermatt, F., Cavender-Bares, J., Carla, G. E., Kentaro, K. S., Hahl, T., et al. (2017). Genomics meets remote sensing in global change studies: monitoring and predicting phenology, evolution and biodiversity. *Curr. Opin. Environ. Sustain.* 29, 177–186. doi: 10.1016/j.cosust.2018.03.005

Yew, C.-L., Kakui, H., and Shimizu, K. K. (2018). Agrobacterium-mediated floral dip transformation of the model polyploid species *Arabidopsis kamchatica*. *J. Plant Res.* 131, 349–358. doi: 10.1007/s10265-017-0982-9

Yoo, M.-J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 October 2020
doi: 10.3389/fgene.2020.567262

A Recently Formed Triploid *Cardamine insueta* Inherits Leaf Vivipary and Submergence Tolerance Traits of Parents

Jianqiang Sun[1†], Rie Shimizu-Inatsugi[2†], Hugo Hofhuis[3], Kentaro Shimizu[4], Angela Hay[3], Kentaro K. Shimizu[2,5] and Jun Sese[6,7*]*

[1] Research Center for Agricultural Information Technology, National Agriculture and Food Research Organization, Tsukuba, Japan, [2] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [3] Max Planck Institute for Plant Breeding Research, Cologne, Germany, [4] Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan, [5] Kihara Institute for Biological Research (KIBR), Yokohama City University, Yokohama, Japan, [6] Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan, [7] Humanome Lab, Inc., Tokyo, Japan

OPEN ACCESS

Edited by:
Yves Van de Peer, Ghent University, Belgium

Reviewed by:
Polina Yu Novikova, Ghent University, Belgium
Jianghua Chen, Xishuangbanna Tropical Botanical Garden (CAS), China

****Correspondence:***
Kentaro K. Shimizu
kentaro.shimizu@ieu.uzh.ch
Jun Sese
sese.jun@aist.go.jp;
sesejun@humanome.jp

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *29 May 2020*
Accepted: *18 August 2020*
Published: *06 October 2020*

Citation:
Sun J, Shimizu-Inatsugi R, Hofhuis H, Shimizu K, Hay A, Shimizu KK and Sese J (2020) A Recently Formed Triploid Cardamine insueta Inherits Leaf Vivipary and Submergence Tolerance Traits of Parents. Front. Genet. 11:567262. doi: 10.3389/fgene.2020.567262

Contemporary speciation provides a unique opportunity to directly observe the traits and environmental responses of a new species. *Cardamine insueta* is an allotriploid species that appeared within the past 150 years in a Swiss village, Urnerboden. In contrast to its two progenitor species, *Cardamine amara* and *Cardamine rivularis* that live in wet and open habitats, respectively, *C. insueta* is found in-between their habitats with temporal water level fluctuation. This triploid species propagates clonally and serves as a triploid bridge to form higher ploidy species. Although niche separation is observed in field studies, the mechanisms underlying the environmental robustness of *C. insueta* are not clear. To characterize responses to a fluctuating environment, we performed a time-course analysis of homeolog gene expression in *C. insueta* in response to submergence treatment. For this purpose, the two parental (*C. amara* and *C. rivularis*) genome sequences were assembled with a reference-guided approach, and homeolog-specific gene expression was quantified using HomeoRoq software. We found that *C. insueta* and *C. rivularis* initiated vegetative propagation by forming ectopic meristems on leaves, while *C. amara* did not. We examined homeolog-specific gene expression of three species at nine time points during the treatment. The genome-wide expression ratio of homeolog pairs was 2:1 over the time-course, consistent with the ploidy number. By searching the genes with high coefficient of variation of expression over time-course transcriptome data, we found many known key transcriptional factors related to meristem development and formation upregulated in both *C. rivularis* and *rivularis*-homeolog of *C. insueta*, but not in *C. amara*. Moreover, some *amara*-homeologs of these genes were also upregulated in the triploid, suggesting *trans*-regulation. In turn, Gene Ontology analysis suggested that the expression pattern of submergence tolerant genes in the triploid was inherited from *C. amara*. These results suggest that the triploid *C. insueta* combined advantageous patterns of parental transcriptomes to contribute to its establishment in a new niche along a water-usage gradient.

Keywords: allopolyploid, homeolog, RNA-seq, meristem formation, ecological niche

INTRODUCTION

The molecular basis of speciation has been a central question in biology (Coyne and Orr, 2004). Little is known still about how a new species obtains new traits to adapt to a distinct environment. A major obstacle in studying this is that most speciation events occurred in the past, and thus the traits and the environment at the time of speciation are not directly observable. The difference in traits and environments between current species may represent evolution after speciation rather than the changes that occurred at speciation. A unique opportunity to study speciation in action is contemporary allopolyploid speciation (Soltis and Soltis, 2009; Abbott et al., 2013). Several cases of polyploid speciation during the past 150 years have been documented, for example in *Tragopogon*, *Senecio*, *Mimulus*, *Spartina*, and *Cardamine* (Urbanska et al., 1997; Abbott and Andrew, 2004; Ainouche et al., 2004; Soltis et al., 2004). Because polyploid speciation immediately confers complete or partial reproductive isolation between the new polyploid and progenitor species, a new polyploid species must establish and propagate while surrounded by individuals with different ploidy. To overcome this situation termed "minor cytotype disadvantage," two traits are suggested to facilitate establishment (Comai, 2005). First, the distinct environmental niche of a polyploid species would reduce competition with progenitor species. Second, clonal vegetative propagation or self-fertilization would assure the persistence of new polyploids at the initial stages because meiotic abnormality is common in newly formed polyploid species (Levin, 2002; Comai, 2005; Cifuentes et al., 2010; Zielinski and Mittelsten Scheid, 2012). This would be critical for odd-ploidy species including triploids, which often contribute to the formation of higher polyploids via a so-called triploid bridge (Bretagnolle and Thompson, 1995; Ramsey and Schemske, 1998; Mable, 2003; Husband, 2004; Tayalé and Parisod, 2013; Mason and Pires, 2015). Despite the significance of these traits, the underlying molecular mechanisms are yet to be studied.

The contemporary polyploid *C. insueta* belongs to the genus *Cardamine*, which has long been studied for ecological polyploid speciation (Howard, 1948; Hussein, 1948), and represents adaptive radiation by recurrent polyploidization along water-usage gradients (Shimizu-Inatsugi et al., 2017; Akiyama et al., 2019). A major advantage to studying *Cardamine* is that it is closely related to the model plant *Arabidopsis thaliana*, and a reference genome assembly of *Cardamine hirsuta* (Gan et al., 2016) is publicly available, thus functional and genomic data of these model species are readily available. One allotriploid species in *Cardamine*, *C. insueta* ($2n = 3x = 24$; RRA), is a textbook example of contemporary speciation discovered by Urbanska-Worytkiewicz and Landolt (1974b). It was formed by the hybridization of two progenitor diploids *Cardamine amara* ($2n = 2x = 16$; AA) as a paternal progenitor and *Cardamine rivularis* ($2n = 2x = 16$; RR, belonging to *Cardamine pratensis* complex sensu lato) as a maternal progenitor approximately 100–150 years ago at the valley of Urnerboden in the Swiss Alps (Urbanska-Worytkiewicz and Landolt, 1972, 1974b; Urbanska et al., 1997; Mandáková et al., 2013; Zozomová-Lihová et al., 2014) (**Figure 1A**). The two diploid progenitors have distinct ecological habitats. While *C. amara* grows in and beside water streams, *C. rivularis* inhabits slightly moist sites, avoiding permeable and fast drying soil (Urbanska-Worytkiewicz and Landolt, 1974a,b) (**Figure 1B**). Around the end of the 19th to the early 20th centuries, the deforestation and land-use conversion to grazing induced the hybridization of these two diploids to produce the triploid species *C. insueta*, which is abundant in manured hay-meadows (Urbanska-Worytkiewicz and Landolt, 1972; Urbanska et al., 1997; Mandáková et al., 2013). Cytogenetic studies suggested that *C. insueta* served as a triploid bridge in the formation of pentaploid and hexaploid *Cardamine schulzii* by the further hybridization with autotetraploid *Cardamine pratensis* (*sensu stricto*, $2n = 2x = 30$; PPPP; hypotetraploid derived from a chromosomal fusion) in Urnerboden (Mandáková et al., 2013).

The propagation of triploids mainly depends on vegetative propagation for two reasons, high male sterility *per se* and hay cutting and grazing in flowering season (Urbanska et al., 1997). One of the progenitor species, *C. rivularis*, can produce plantlets on the surface of leaves and nodes by ectopic meristem formation, which is a common feature of the *C. pratensis* complex (Smith, 1825; Salisbury, 1965; Dickinson, 1978). This characteristic is inherited by *C. insueta*, enabling it to be a dominant species at the site despite its ploidy level (Urbanska-Worytkiewicz and Landolt, 1974a; Urbanska et al., 1997). This type of leaf vivipary is only found in a limited number of angiosperms and assumed to contribute to population establishment in polyploids (Dickinson, 1978). In this sense, the trait of leaf vivipary can be considered a key factor for the establishment of this triploid.

Another interesting aspect of *C. insueta* establishment is its ecological niche shift relative to its progenitor species. Genus *Cardamine* is known to include many submergence tolerant species including *C. amara* (Shimizu-Inatsugi et al., 2017; Akiyama et al., 2019). An allotetraploid *Cardamine flexuosa*, derived from *C. amara* and *C. hirsuta* diploid progenitors, was shown to inherit parental traits and be successful in a wider soil moisture range (Shimizu-Inatsugi et al., 2017; Akiyama et al., 2019). The transcriptomic response of *C. flexuosa* to submergence or drought stress was shown to be combined although attenuated compared to its progenitor species, which could confer the wider tolerance found in the polyploid. Even though the niche separation between *C. rivularis* and *C. insueta* is not yet clearly illustrated, our field observations are consistent with this hypothesis.

In this study, we focused on the time-course gene expression pattern of the triploid *C. insueta* and its two diploid progenitors during submergence treatment, which induces both water stress and ectopic meristem formation on leaves. To study the time-course data of homeologs, we employed bioinformatic methods of variably expressed genes because data points of a time-course are not independent and serve partly as replicates (Yamaguchi et al., 2008; Shin et al., 2014). Here we combined the time course analysis with subgenome-classification bioinformatic workflow of HomeoRoq (Akama et al., 2014), and detected variably

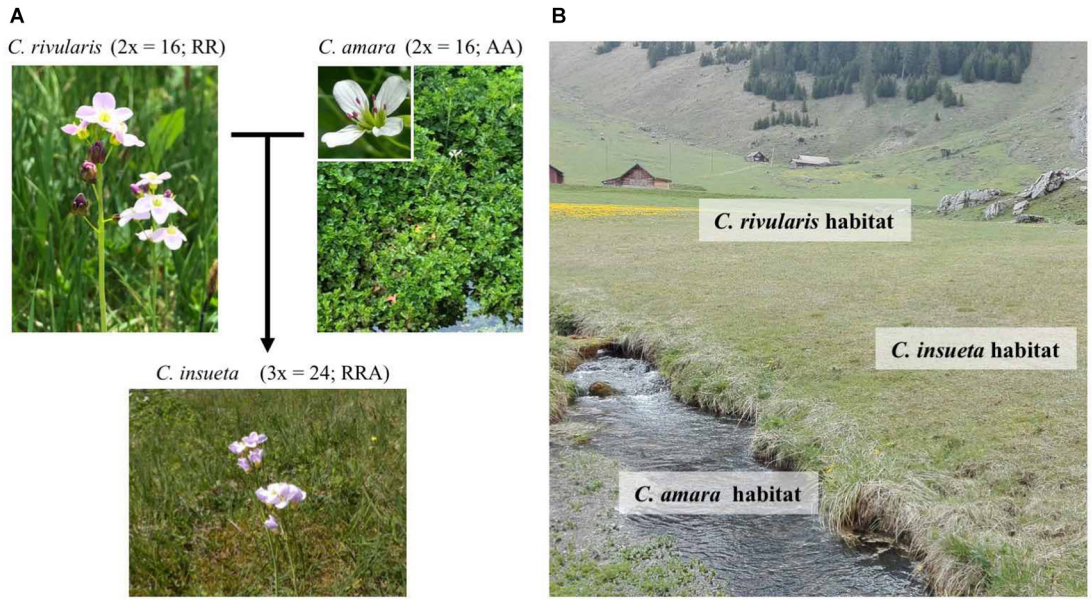


FIGURE 1 | Habitats and relations among the three *Cardamine* species. **(A)** *C. insueta* (triploid) was naturally formed by hybridization between *C. amara* (diploid) and *C. rivularis* (diploid) 100–150 years ago in their natural habitat in Swiss Alps. **(B)** Conceptual indication of the habitats of three species at Urnerboden: *C. amara* prefers wet habitats along waterside; *C. rivularis* prefers meadow; and allotriploid *C. insueta* can be found between them.

expressed homeologs (VEH) during the treatment. We address the following specific questions:

(1) What is the expression level and the ratio of homeologous genes in triploid species in response to submergence, either genome-wide or between each homeologous gene pair?
(2) Which kind of genes are enriched in VEH? Do they reflect the phenotypic trait of each progenitor species or the triploid? How does *C. insueta* combine the expression patterns of the two progenitors?

MATERIALS AND METHODS

Plant Materials and RNA Sequencing

Cardamine insueta, *C. amara*, and *C. rivularis* plants used in this study were collected from Urnerboden. All plants were grown together in a plant cultivation room with 16 h light and 8 h dark cycle. The plants were planted in single pots, placed on trays, and watered from below.

Submergence treatment was started in the morning at 07:00. Two mature leaves were detached and submerged in water. We isolated RNA from the floating leaflets of the three species at nine time points after the start of submergence treatment (0, 2, 4, 8, 12, 24, 48, 72, and 96 h) using Qiagen RNeasy kit (Qiagen, Maryland, United States). RNA quality was assessed by Bioanalyser Nanochip (Agilent, Santa Clara, United States) and libraries quantified by Qubit (Thermo Fisher, Waltham, MA, United States). In total 27 libraries (3 species × 9 time points) were prepared according to NEBNext Ultra^{textTM} Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, United States) followed by paired end sequencing (100 bp × 2) on a HiSeq2000 with a HiSeq Paired-End Cluster Generation Kit and HiSeq Sequencing Kit (Illumina, San Diego, CA, United States). Trimmomatic (ver. 0.36) (Bolger et al., 2014) was used for discarding the low-quality reads with parameters of "PE -threads 4 -phred33 ILLUMINACLIP:adapters.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50".

Reference Sequence Assembly

The reference sequences of *C. amara* genome (A-genome) and *C. rivularis* genome (R-genome) were assembled by single-nucleotide polymorphism (SNP) substitution at coding regions from the *C. hirsuta* genome (H-genome) (Gan et al., 2016) with the following steps. To assemble the reference sequence of A-genome, first, we pooled all RNA-Seq reads of the nine RNA-Seq samples of *C. amara*. Second, we mapped the reads onto the reference sequence (i.e., H-genome) using STAR (ver. 2.3.0e) (Dobin et al., 2013). Third, we detected SNPs and short indels from the mapping result using samtools (ver. 0.1.18) (Li et al., 2009). SNPs and indels were defined as the polymorphic loci where at least 80% of reads have the alternative nucleotides. Fourth, we replaced the nucleotides on the reference with the alternative nucleotides, if the alternative nucleotide was covered by at least five reads. Finally, the gene annotations of the assembled sequence were converted from the H-genome annotations with the replacement information. To improve the accuracy of sequence, we used the assembled sequence as a reference sequence, and repeated steps two through five, nine times. The resulting A-genome was used for the mapping of individual RNA-seq data from all three species. The R-genome was also reconstructed with the same protocol. As a result,

1,496,561 and 1,484,186 SNP regions on the H-genome were replaced for A-genome and R-genome, respectively.

Evaluation of HomeoRoq Classification Confidence Using Diploids

We used HomeoRoq (ver. 2.1) (Akama et al., 2014) to classify genomic origins of homeolog-specific reads in the nine *C. amara* and *C. rivularis* samples. Following the HomeoRoq pipeline, for each *C. amara* sample, we used STAR to map reads onto the A-genome and R-genome and used HomeoRoq to classify reads as *A-origin*, *R-origin*, and *unclassified*. Then, we calculated the percentage of misclassified reads (i.e., the reads that were classified as *R-origin*). Similarly, we used HomeoRoq to calculate the percentage of misclassified reads (i.e., the reads that were classified as *A-origin*) in each *C. rivularis* sample.

Homeolog Expression Quantification and A-Origin Ratio Definition of Triploid

We used HomeoRoq to analyze the nine *C. insueta* samples. For each *C. insueta* sample, we used STAR to map reads onto A-genome and R-genome and used HomeoRoq to classify reads as *A-origin*, *R-origin*, and *unclassified*. Then, we customized HTSeq (Planet et al., 2012) to count the number of read pairs that mapped on homeolog region for *A-origin*, *R-origin*, and *unclassified* reads of each *C. insueta* sample separately. In the customized HTSeq, if a read mapped on the region overlapped by multiple homeologs, a read was divided by the number of homeologs.

To calculate the number of fragments per kilobase mapped (FPKM) for *C. insueta* samples (I_A and I_R samples), we first allocated the *unclassified* reads into *A-origin* and *R-origin* reads with A-origin ratio. A-origin ratio of homeolog h at the time point S was defined as $p_h^s = a_h^s/(a_h^s + r_h^s)$, where a_h^s and r_h^s are the numbers of *A-origin* and *R-origin* reads of homeolog h at the time point S, respectively. Thus, the number of *A-origin* reads after *unclassified* reads allocation ($a_h'^s$) was calculated as $a_h'^s = a_h^s + u_h^s p_h^s$, where u_h^s is the number of *unclassified* reads of homeolog h in sample S. Similarly, $r_h'^s = a_h^s + u_h^s(1 - p_h^s)$ for *R-origin* reads. Then, FPKM of *A-origin* reads of homeolog h in sample S was calculated as $10^9 a_h'^s/(L_h^A A^s)$, where L_h^A is the length of homeolog h on A-genome and A^s is the total number of *A-origin* reads in sample S; likewise, FPKM of *R-origin* reads was calculated as $10^9 a_h'^s/(L_h^R R^s)$, where L_h^R is the length of homeolog h on R-genome and R^s is the total number of *R-origin* reads in sample S.

In addition, FPKM of progenitors were calculated from the total number of reads (i.e., $a_h^s + u_h^s + r_h^s$). Therefore, FPKM of *C. amara* and *C. rivularis* were calculated as $10^9(a_h^s + u_h^s + s_h^s)/(L_h^A A^s)$ and $10^9(a_h^s + u_h^s + s_h^s)/(L_h^A R^s)$, respectively.

Expressed Homeologs and PCA Analysis

An expressed homeolog was defined as a homeolog with FPKM > 1.0. A homeolog expressed in a sample [i.e., either *amara*-derived in *C. insueta* (I_A), *rivularis*-derived in *C. rivularis* (I_R), *C. amara* or *C. rivularis*] was defined as a homeolog with FPKM > 1.0 at least at one of the nine time points. In total, 21,131 homeologs were expressed at least in one sample. PCA was performed against $\log_{10}$-transformed FPKM of these 21,131 expressed homeologs. To avoid calculating $\log_{10}0$, the $\log_{10}$-transfromed FPKM was truly calculated as $\log_{10}(\text{FPKM} + 1)$.

Identification of Variably Expressed Homeologs (VEH) and Gene Ontology (GO) Enrichment Analysis

Mean and coefficient of variation (CV) were calculated from $\log_{10}$-transformed FPKM over the nine time points. VHE was defined as an homeolog satisfied the mean > 1.0 and the CV > 0.20. We identified from I_A, I_R, *C. amara*, and *C. rivularis* samples, separately.

Gene ontology (GO) enrichment analysis was performed for the four variably expressed homeolog (VEH) sets with R packages clusterProfiler (ver. 3.12.0) and org.At.tair.db (ver. 3.8.2) (Yu et al., 2012). To remove redundancies of GO categories, only GO categories which are associated with 10–500 *Cardamine* homeologs and below the third level in the GO category hierarchy were used. The threshold FDR = 0.1 was used for cutoff of significantly enriched GO categories.

RESULTS

Plantlet Induction on *C. insueta* and *C. rivularis* Leaves by Submergence

At the field of Urnerboden valley, we could scarcely observe normal seed setting on *C. insueta*, but small plantlets on leaves were frequently observed after flowering, as described previously (Urbanska, 1977). We also observed small plantlets on the leaves of *C. rivularis*. In contrast, *C. amara* does not form plantlets on leaves, rather adventitious roots and shoots were formed from rhizomes. In the natural habitat, the plantlet formation of *C. rivularis* and *C. insueta* can be seen at flowering to post-flowering season (Salisbury, 1965; Urbanska, 1977). It was also reported that *C. pratensis* (which is closely related to *C. insueta* or considered the same species) tend to bear more plantlets on the leaves in damper sites than in drier sites (Salisbury, 1965), implying that high moisture could be the trigger for meristem formation. Thus, we tested plantlet induction by submergence treatment using dissected leaves with this trio of species in the lab. We detached mature leaves from mother plants propagated in a climate chamber and floated the leaves on water. Within 16 h, we observed the activation of dormant shoot meristems and initiation of ectopic root meristems, which formed visible plantlets on *C. rivularis* leaves 96 h after submergence (**Supplementary Figure S1** and **Supplementary Dataset S1**). Induction of ectopic plantlets followed a similar time-course in *C. insueta*. In contrast, plantlet induction was not observed on the leaves of *C. amara*. In addition, during the 96-h treatment, no symptoms of necrosis appeared on any of the leaves, suggesting that all three species have some submergence tolerance for at least 96 h.

Gene Annotation on the Two Diploid Progenitor Reference Sequences

To detect how homeologous genes are expressed in plantlet induction and submergence treatment, we harvested time-course RNA-Seq samples of *C. insueta* and diploid progenitor leaves at nine time points after initial submergence (i.e., 0, 2, 4, 8, 12, 24, 48, 72, and 96 h) (**Supplementary Figure S2A**). We harvested the first lateral leaflet pair in young leaves with no ectopic plantlets. To quantify homeolog-specific gene expression, we assembled the genomes of *C. amara* (A-genome) and *C. rivularis* (R-genome), respectively, using the same pipeline of a reference-guided approach using RNA-Seq reads (**Supplementary Figure S2B**). The genome sequence of a close relative, *C. hirsuta* (H-genome) (Gan et al., 2016), was used as a reference. The A-genome structure is reported to be almost perfectly collinear with that of H-genome, except for one pericentric inversion at chromosome 1, by cytological studies (Mandáková et al., 2013, 2014). The genome structures of the A-genome and R-genome are also similar to each other (Mandáková et al., 2013). The length of assembled reference sequences of A-genome and R-genome are 198,651,635 and 198,654,862 nucleotides, respectively, which are nearly the same as the length of the original H-genome (198,654,690 nucleotides). We also annotated the orthologous genes of *C. amara* and *C. rivularis* according to the information of *C. hirsuta* H-genome. In total, we found 23,995 and 24,115 genes covered by at least one read among the nine time points on the assembled A-genome and R-genome, respectively. These gene sets, which correspond to 81.5 and 81.7% of 29,458 genes in H-genome, respectively, were defined as expressed and used for the following analysis.

Expression Ratio From Each Subgenome Is Consistent With the Number of Chromosomes

We applied the HomeoRoq analysis pipeline (Akama et al., 2014) to map RNA-Seq reads of *C. insueta* samples to A-genome and R-genome, and classify the origin of each RNA-seq read of *C. insueta* samples to either *A-origin* (i.e., the genomic origin of the read is A-subgenome) or *R-origin* (**Supplementary Figure S2C**). After filtering for read quality, 10.6 million read pairs on average among the nine samples could be classified as homeolog-specific read pairs (**Supplementary Dataset S2**). Of the total homeolog-specific read pairs in the *C. insueta* 0 h sample, 27.3 and 56.7% of read pairs were classified as *A-origin* and *R-origin*, respectively. To confirm that *A-origin* and *R-origin* reads were correctly mapped to A-genome and R-genome, respectively, we checked the alignments of several highly expressed homeologs in the mapping results with Integrative Genomics Viewer (IGV) (**Supplementary Figure S3**) (Robinson et al., 2011). We found only a few SNPs in the alignments between *A-origin* reads and A-genome, and *R-origin* reads and R-genome, respectively. Considering the young origin of *C. insueta* within 150 years, we can assume that the genomic distances between *C. insueta* and its progenitors are very small, and HomeoRoq can manage this range of difference with low error rate (Kuo et al., 2020). Besides *A-origin* and *R-origin* reads, the remaining 16.0% of read pairs could be classified to neither *A-origin* nor *R-origin* (*unclassified*) due to the lack of SNPs or the identical sequence on the correspondence region. As a whole genome, the ratio of *A-origin* to *R-origin* reads was approximately 1:2.

When we analyzed all samples from the other eight time points, we observed a slight increase in the proportion of *A*-origin reads in correlation with the time point, from 1:2.07 at 0 h to 1:1.90 at 96 h (**Supplementary Dataset S2**). Instead of this minor transition, the expression ratio between subgenomes remained A:R $\approx$ 1:2 with *C. insueta* samples at all time points, indicating that the expression ratio from each subgenome is consistent with the number of chromosome regardless of the submergence treatment.

Most Homeolog Pairs Were Expressed in Proportion to the Subgenomes in *C. insueta*

To investigate the proportion of expression levels of homeolog pairs in *C. insueta*, we quantified the expression level of each homeolog pair at each time point. We found that (i) the correlation between the expression levels of homeolog pairs was higher than 0.81 at any time point (**Figure 2A** and **Supplementary Figure S4**). However, (ii) the expression levels of most homeologs expressed from the A-subgenome (A-homeolog) were approximately half that of R-homeologs. To understand the proportion of expression levels of homeolog pairs in detail, we calculated A-origin ratio—the proportion of A-homeolog expression level to the total A-homeolog and R-homeolog expression levels—for all homeolog pairs at each time point. We found that the distribution of A-origin ratios had a gentle peak at the position of 0.33 at all time points (**Figure 2B** and **Supplementary Figure S5**). This result suggests the expression ratio of the majority of homeolog pairs is consistent with the copy number, i.e., the subgenome-set numbers of the triploid. In addition, we found two sharp peaks at both edges, the positions of 0.0 and 1.0, of A-origin ratio, which represent the homeologs only expressed in either subgenome.

Additionally, to investigate whether the A-origin ratio changes during the submergence treatment, we compared the A-origin ratio distributions between different time points. The patterns of all time points were correlated to each other, with the least coefficiency (0.66) between 0 and 2 h (**Figure 2B** and **Supplementary Table S1**). This result indicates that A-origin ratios did not change drastically in most homeolog pairs by the submergence treatment, but a limited number of homeolog pairs change the expression balance.

The Whole Genome Expression Pattern of Each *C. insueta* Subgenome Is Closer to That of Its Progenitor Genome

To gain an overview of how homeologous gene expression varies at the whole genome level among *C. insueta* and the progenitor species *C. amara* and *C. rivularis*, we conducted principal component analysis (PCA). PCA was performed against the $\log_{10}$-transformed FPKM of 21,131 expressed homeologs (**Figure 3**). We found that the first principal component (PC1)

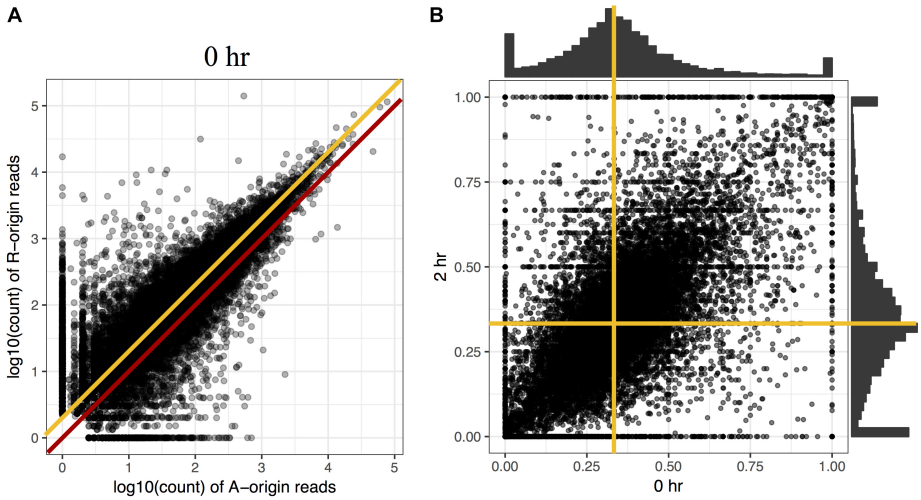


FIGURE 2 | Comparison of homeolog expression from A- and R-subgenomes. **(A)** Expression ratio between A- and R-homeologs before submergence treatment in the triploid *C. insueta*. Each dot shows the relation between the log10-transformed A-origin and R-origin read of a homeolog pair at 0 h point. Only the homeolog pairs with FPKM > 1.0 in either I_A or I_R samples are shown. The red line represents the ratio *A*:*R* = 1:1, and the orange line represents the ratio *A*:*R* = 1:2. **(B)** Comparison of A-origin ratios between two time points, 0 and 2 h, in the triploid *C. insueta*. Each point shows the A-origin ratios of a homeolog pair at 0 and 2 h. The orange lines represent the position of *A*-origin = 0.33 at each time point.

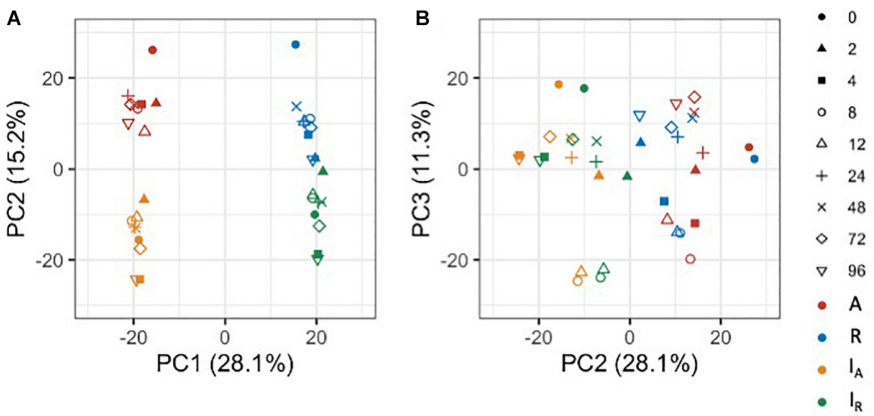


FIGURE 3 | Principal component analysis of the expressed homeologs/genes in *C. insueta* (I_A), *C. insueta* (I_R), *C. amara* (A) and *C. rivularis* (R) samples at 9 time points. PCA was performed against $\log_{10}$-transformed FPKM of 21,131 expressed homeologs. The two plots show the relation between PC1-PC2 **(A)** and PC2-PC3 **(B)**. The colors represent genome/subgenome, and the shapes represent the time points after the start of submergence treatment.

grouped samples into two groups: the one with A-homeologs of *C. insueta* (I_A) and *C. amara* (*A*) samples and the other with R-homeologs of *C. insueta* (I_R) and *C. rivularis* (*R*) samples. In addition, we also found that the second principal component (PC2) grouped samples into two groups: one consisting of polyploid samples (I_A and I_R samples, lower side of **Figure 3A**) and the other consisting of diploid samples (*A* and *R* samples, upper side of **Figure 3A**). By PC1 and PC2, the samples were grouped into four clusters according to the subgenome type. In contrast, by PC2 and the third principal component (PC3), we observed the transition according to the treatment time, showing a characteristic transition from 0 to 12 h, and the recurrence of 24, 48, 72, and 96 h samples toward 0 h samples in each subgenome, which might reflect the combined effect of submergence stress and circadian rhythm (**Figure 3B**). The result of PC1 suggests that the majority of the homeologs of I_A and I_R should retain a similar expression pattern to each parent, *A* and *R*. When we focus on PC2, the distance between *R* and I_R is slightly closer than that between *A* and I_A. This might reflect the difference in the number of subgenome sets in the triploid, *A*:*R* = 1:2, implying a stronger effect from the progenitor with more subgenome sets.

VEHs Related to Submergence and Their GO Enrichment Analysis

To understand the difference among species in plantlet formation on the leaf and in submergence response, we focused on the homeologs with a higher expression change during the

treatment. Standard tools to identify differentially expressed genes between different conditions are not directly applicable to time-course data, in which expression levels of neighboring time points may be highly correlated. We defined variably expressed homeologs (VEHs) according to the coefficient of variation (CV) among the expression levels of the nine time points, since CV is used for identifying variably expressed genes in various studies involving time-course analysis (Czechowski et al., 2005; Yamaguchi et al., 2008; Shin et al., 2014; Zhao et al., 2017). We identified 1,194, 1,144, 1,030, and 1,063 VEHs from I_A, I_R, A, and R genome/subgenome with the cutoff CV > 0.2 throughout the treatment, respectively (**Supplementary Dataset S3**). We visualized the patterns by focusing on two genes that were expected to be affected (**Supplementary Figure S6**). The genes associated with ethylene-response such as *ERF1* (AT3G23240) (Chao et al., 1997; Solano et al., 1998) and circadian rhythm such as *CCA1* (AT2G46830) (Alabadí et al., 2001) were identified as VEHs in all samples, which should reflect the ethylene-response to submergence and circadian rhythm response, respectively. The expression pattern of these two homeologs were similar among all four VEH sets from I_A, I_R, A, and R (**Supplementary Figure S6**). In addition to these common VEHs, we also found more homeologs identified as VEHs only in one to three samples (**Supplementary Figure S7**).

To investigate the biological processes of VEH sets of I_A, I_R, A, and R, we performed gene ontology (GO) enrichment analysis against the four VEH sets (**Table 1** and **Supplementary Dataset S4**). The numbers of enriched GO categories were 146, 155, 160, and 181, respectively for I_A, I_R, A, and R. A value of negative log10(q-value) more than 1 was defined as significant, and a higher value indicates stronger enrichment. We found that some of the GO categories related to water stress, including GO:0006066 (alcohol metabolic process), GO:0009723 (response to ethylene) and GO:0009414 (response to water deprivation), were enriched in all four VEH sets (**Table 1**). As gas diffusion rates are restricted under water, submergence of plants induces ethylene accumulation and low oxygen availability, which could result in the reorganization of the ethylene-response pathway and fermentation pathway (e.g., anaerobic respiration and alcohol metabolism). The enriched categories GO:0006066 and GO:0009723 indicate that I_A, I_R, A, and R all respond to ethylene and hypoxia signals with the submergence treatment. Two alcohol related categories (GO:0006066 alcohol metabolic process and GO:0046165 alcohol biosynthetic process) were more strongly enriched in A and I_A, which was two orders of magnitude higher than I_R and R [>2 difference in negative log10(q-value) in **Table 1**]. In addition, GO:0009414 (response to water deprivation), which encompasses the expression changes of aquaporin genes and ethylene-responsive genes (**Supplementary Dataset S5**), was enriched.

In contrast, some GO categories related to submergence stress were only above the significance threshold in part of the four VEH sets with various combinations (**Table 1**). The two categories related to ethylene metabolism, GO:0009873 (ethylene-activated signaling pathway) and GO:0071369 (cellular response to ethylene stimulus), were not detected in I_R but all other three. All these ethylene related GO categories were most strongly enriched in A, suggesting larger number of genes are detected than other VEH sets. In addition, the categories related to abscisic acid signaling, which is known to work antagonistically to ethylene, GO:0009738 (abscisic acid-activated signaling pathway) and GO:0071215 (cellular response to abscisic acid stimulus), were also detected only in A with many inactivated genes by treatment. In contrast, the categories related to oxidative stress showed the strongest enrichment in R than others, suggesting higher intensity of oxidative stress in *C. rivularis* than other species.

TABLE 1 | The negative log10(q-value) of the enriched GO categories in each VEH set described in the manuscript.

Keyword	Accession	VEH set				GO name
		A	I_A	I_R	*R*	
Alcohol	GO:0006066	6.9	5.2	2.7	2.0	Alcohol metabolic process
	GO:0046165	6.4	6.4	4.4	2.7	Alcohol biosynthetic process
Water	GO:0009414	3.9	3.3	5.1	7.3	Response to water deprivation
Ethylene	GO:0009723	7.9	3.6	2.8	7.5	Response to ethylene
	GO:0009873	6.1	15	ND	2.5	Ethylene-activated signaling pathway
	GO:0071369	6.1	1.3	ND	2.3	Cellular response to ethylene stimulus
	GO:0009692	2.6	ND	ND	ND	Ethylene metabolic process
	GO:0009693	2.6	ND	ND	ND	Ethylene biosynthetic process
Abscisic acid	GO:0009738	1.4	ND	ND	ND	Abscisic acid-activated signaling pathway
	GO:0071215	1.4	ND	ND	ND	Cellular response to abscisic acid stimulus
Oxidative stress	GO:0006979	19	ND	ND	2.6	Response to oxidative stress
	GO:0042743	ND	ND	ND	3.0	Hydrogen peroxide metabolic process
	GO:2000377	ND	1.6	1.5	5.6	Regulation of reactive oxygen species metabolic process
	GO:0010310	ND	1.3	1.1	5.1	Regulation of hydrogen peroxide metabolic process
Meristem	GO:0035266	ND	3.1	1.9	ND	Meristem growth
	GO:0010075	ND	3.4	2.0	ND	Regulation of meristem growth
	GO:0048509	ND	2.8	1.8	ND	Regulation of meristem development

*This data is the extract from the list of all enriched GOs (**Supplementary Dataset S4**). ND means that the category was not detected as enriched by the threshold FDR = 0.1.*

VEHs Related to Meristem and Their GO Enrichment Analysis

Among the GO categories enriched in four VEH sets, three categories were related to meristem activity: GO:0035266 (meristem growth), GO:0010075 (regulation of meristem growth), and GO:0048509 (meristem development) (**Table 1**). They were enriched only in VEH sets of I_A and I_R, but

not in *A* and *R*, although *C. rivularis* can also produce ectopic meristems.

We analyzed the expression pattern of several known transcriptional factors which could be involved in ectopic meristem formation and development in *Cardamine* (**Figure 4** and **Supplementary Figure S8**). Class I Knotted1-like homeobox (KNOX) transcription factors function to maintain shoot apical meristem activity in many different plant species (Vollbrecht et al., 1991; Long et al., 1996; Hay and Tsiantis, 2010). Importantly, the overexpression of *SHOOTMERISTEMLESS* (*STM*) and another KNOX gene, *Arabidopsis knotted 1*-like gene (*KNAT1*), are known to cause ectopic meristem formation on the leaf in *A. thaliana* (Chuck et al., 1996; Williams, 1998). Moreover, an *STM* ortholog is required for leaf vivipary in *Kalanchoë daigremontiana* (Garcês et al., 2007), a clonal propagation trait that is also observed in *C. rivularis* and *C. insueta*. As summarized in **Figure 4**, orthologs of the four *A. thaliana* KNOXI genes, *STM*, *KNAT1*, *KNAT2 and KNAT6*, showed upregulated expression in all or some of I_A, I_R and *R*, but not in *A*. In addition, we also found that *PDF1* increased expression levels in *C. insueta* (both I_A and I_R) and *R*, which is exclusively detected in the L1 layer of shoot apical meristem throughout the shoot development of *Arabidopsis* (Abe et al., 1999). Three other transcription factor-encoding genes, *CUC2*, *CUC3*, and *LAS*, which contribute to ectopic shoot apical meristem in tomato leaves (Rossmann et al., 2015), were induced in *R* and I_R but not in *A*.

The expression of genes related to root apical meristem maintenance and formation showed similar patterns to those related to shoot apical meristem formation. Transcription factors with an AP2/ERF domain for the maintenance of root apical meristem [*PLT1*, *PLT2*, and *PLT3*; (Drisch and Stahl, 2015)], were scarcely expressed in *A* but induced in others. For genes responsible for radial patterning, *SHR* and *SCR*, expression level of *SHR* was increased in all four sets, while that of *SCR* was upregulated only temporarily and reverted in 24 h. The expression of *WOX5*, a key factor to maintain the root stem cell (Sarkar et al., 2007), was very low in all sets, most probably due to the extremely limited expression area only at the quiescent center.

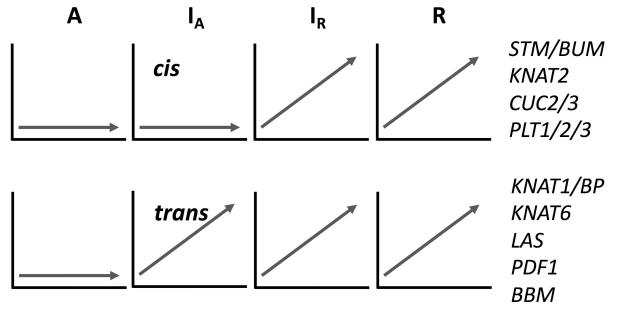


FIGURE 4 | A schematic drawing of *cis*- and *trans*-regulation of key regulatory genes in meristem formation. The expression patterns of I_A homeologs are arbitrary categorized to *cis*- or *trans*-regulated according to the expression pattern of each homeolog. See **Supplementary Figure S8** for their original temporal expression patterns.

Many of the above-mentioned transcription factors contributing to meristem formation and maintenance are known to be related to or controlled by auxin, thus the transportation of auxin might be also involved in ectopic meristem development in *C. insueta* and *C. rivularis*. One of the auxin transporter genes, *PIN1*, was induced by the treatment in all four sets soon after the start of submergence, but after 24 h the high expression level was only retained in I_A, I_R, and *R*. On the other hand, the other auxin transporter genes *PIN3*, *PIN4*, and *PIN7* were temporarily induced by the treatment, but soon decreased among all sets. The peaks of expression of these three genes were at 4–12 h in *A*, 12–24 h in *R*, and 8–24 h in I_A and I_R, suggesting involvement in meristem formation and development. By the VEH enrichment analysis, GO:0060918 (auxin transport) and GO:0009926 (auxin polar transport) were enriched in I_A and I_R. In contrast, the other two auxin related GO categories, GO:0009850 (auxin metabolic process) and GO:0009851 (auxin biosynthetic process), were enriched in only two parents, *A* and *R*.

DISCUSSION

Applicability of HomeoRoq to Diverse Ploidy Levels

HomeoRoq was developed to classify genomic origins of RNA-Seq reads of allopolyploids consisting of two subgenomes (Akama et al., 2014), and has already been applied to *Arabidopsis kamchatica* ($2n = 4x = 32$; HHLL), an allotetraploid between two diploids of *Arabidopsis halleri* ($2n = 2x = 16$; HH) and *Arabidopsis lyrata* ($2n = 2x = 16$; LL). Here, we successfully applied HomeoRoq to another species with a different ploidy level. The average proportions of the reads mapped on the wrong genome in *C. amara* and *C. rivularis* samples were $1.1 \pm 0.1\%$ and $1.2 \pm 0.1\%$, respectively (**Supplementary Dataset S2**). This high accuracy is comparable to the evaluation of the *A. kamchatica* data, 1.23–1.64% (Akama et al., 2014; Kuo et al., 2020).

The proportion of *unclassified* reads in this study, which has the same matching rates on both parental genomes, was very close to that in the *A. kamchatica* study. In this study, $11.5 \pm 2.0\%$ of reads in *C. insueta* samples were *unclassified* on average, compared to 11.0% in *A. kamchatica* (Akama et al., 2014), suggesting a similar divergence level between subgenomes in the two cases. Considering the percentage of unclassified reads and the low misclassification rate with diploid progenitors, HomeoRoq can be applied to genomes of any ploidy level providing that the genome consists of two types of subgenome.

Total Gene Expression Level of Each Subgenome Is Consistent With the Chromosome Number

The ratio of *A-origin* to *R-origin* reads in *C. insueta* was approximately 1:2. This result is consistent with the distribution of A-origin ratio showing a gentle peak at around 0.33 with a smooth decrease toward the edges (**Figure 2**). This distribution indicates that expression ratios of most homeologs correlates with the copy number. A similar tendency could be found in

other Brassicaceae allotetraploids (Akama et al., 2014; Douglas et al., 2015; Yang et al., 2016). In the analysis of triploid banana ($2n = 3x = 33$; ABB), a hybrid between *Musa acuminata* ($2n = 2x = 22$; AA) and *Musa balbisiana* ($2n = 2x = 22$; BB), the read proportion is distributed around 0.66 for the B alleles by 155 homeologs with rather high expression level detected by LC-MSMS as isoforms (van Wesemael et al., 2018). This could also be seen in hexaploid bread wheat consisting of three subgenomes, where 70% of genes showed balanced expression among homeologs (Ramírez-González et al., 2018). So far, this consistency between ploidy number and expression ratio looks like a general rule in many species with some exceptions like tetraploid cotton (Yoo et al., 2013).

In addition to the majority of genes that show balanced expression between homeologs, a limited proportion of genes show significant differential expression. Even though a direct comparison among studies is difficult due to different thresholding policies, the number of genes with unbalanced homeolog expression tends to be the minor fraction in many quantitative studies. Further studies should show whether a similar pattern is observed in even higher ploidy levels or other odd ploidies.

Limited Number of Homeolog Pairs Changed Expression Ratio in Submergence Condition

Though the number of homeologs with unbalanced expression is smaller than that with balanced expression, they could play a significant role in speciation of polyploid species, especially for achieving a combined trait from progenitors. A series of studies have reported that homeolog expression ratios can be changed depending on external environments (Bardil et al., 2011; Dong and Adams, 2011; Akama et al., 2014; Paape et al., 2016). Akama et al. (2014) evaluated the changes of the homeolog expression ratio of *A. kamchatica* after cold treatment. They reported that the homeolog expression ratios before and after cold treatment were highly correlated ($R^2 = 0.87$), and only 1.11% of homeolog pairs statistically significantly changed in expression ratios in response to cold treatment (Akama et al., 2014). A similar result was reported for zinc treatment of *A. kamchatica.* The correlation of homeolog expression ratios between zinc treatment and control ranged from 0.89 to 0.94, and 0.3–1.5% of homeologs significantly changed expression ratios after Zn treatment (Paape et al., 2016).

In this study using another Brassicaceae species, *C. insueta*, the correlation coefficients of A-origin ratios between 0 h and the other time-points ranged from 0.68 to 0.82 (**Supplementary Table S1**). The lowest correlation occurring between 2 h and other time points may suggest that the initial reaction to the treatment had the strongest effect on gene expression. The overall high correlations among time points indicate that the expression ratios of most homeologs do not change considerably in response to treatment. Even though *C. rivularis* and *C. amara* show species-specific responses to submergence, leaf vivipary and submergence tolerance, respectively, no specific expression preference or dominance of either progenitor was detected in the triploid. This suggests that transcriptional changes in only a limited number of homeologs, rather than genome-wide, might be responsible for the control of physiological change under submergence conditions.

Triploid Inherited Advantageous Traits From Progenitors

Only about 6% of the expressed genes were detected as VEH throughout the 96-hr treatment in each genome and subgenome, suggesting the criteria were fairly conservative. Among enriched GO categories are water stress related ones, particularly ethylene-response and fermentation. Fermentation metabolism in plants is important for submergence stress. We found more VEH genes in the fermentation-related categories in the diploid *C. amara* and the *amara*-derived subgenome of *C. insueta* than counterparts (**Table 1**). This suggests that *C. insueta* inherited the fermentation ability as a submergence response more largely from *C. amara* side. The ethylene signaling pathway should have been stimulated in all three species as many related GO categories are found enriched in all VEH genes. However, the stress level seems to be variable according to the species as shown in the difference of enriched GO categories. In all of these ethylene related GO categories, *C. amara* had the strongest enrichment (i.e., highest number of VEH genes), and the enrichment in *amara*-derived subgenome was stronger than in *rivularis*-derived subgenome in *C. insueta*. These enrichment intensities should suggest that *C. amara* has higher acclimation ability to submergence through an activation of alcohol metabolic pathway and alteration in hormone signaling pathway and thus suffer from less oxidative stress as a result, as speculated by its habitat and a previous study (Shimizu-Inatsugi et al., 2017). In addition, in *C. insueta*, the contribution to the stress response of I_A seems larger than that of I_R, found as stronger enrichment in I_A than in I_R.

GO enrichment analysis with VEH genes also showed three GO categories related to meristem, GO:0035266 (meristem growth), GO:0010075 (regulation of meristem growth) and GO:0048509 (meristem development). They were only enriched in the VEH sets of I_A and I_R, but not above the significance threshold in two parents, despite the fact that *C. rivularis* also produces plantlets on the leaf by the activation of ectopic meristems. This might imply that the ability to form ectopic plantlets in response to submergence is enhanced in the triploid *C. insueta* compared to the diploid *C. rivularis*. Considering the disadvantage in sexual reproduction due to the odd ploidy, effective vegetative propagation through plantlets might have been critically important for *C. insueta*.

The expression pattern of known key regulatory genes that function to maintain meristem activity showed two typical patterns, as shown in **Figure 4** and **Supplementary Figure S8**, although mixed patterns are also found. Expression of both types of genes was upregulated in *C. rivularis* (*R*, **Figure 4**) but not in *C. amara* (*A*, **Figure 4**) in response to submergence. Expression of these genes was also upregulated in the *C. insueta* subgenome I_R, but followed two different patterns in the I_A subgenome. These patterns could be categorized as either non-induced, similar to *C. amara*, or induced, similar to *C. rivularis*, suggesting that non-induced homeologs could be *cis*-regulated

(i.e., difference in the *cis*-regulatory regions derived from the two progenitors), and induced homeologs could be *trans*-regulated by I_R. One possibility is that this difference reflects the developmental timing of gene expression during meristem formation and the divergence of *cis*-regulation. For example, the *cis*-regulated genes *STM* and *CUC2* are expressed earlier during embryogenesis in *A. thaliana* than the *trans*-regulated genes *KNAT6* and *KNAT1/BP* (Hay and Tsiantis, 2010). This variation might imply a regulatory relationship among these genes in the gene regulatory network controlling plantlet formation in *C. insueta* leaves. This type of information might provide insights that warrant further study into the molecular mechanism of leaf vivipary in *C. rivularis* and *C. insueta*.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in DNA Data Bank of Japan (DDBJ) Sequence Read Archive (DRA), www.ddbj.nig.ac.jp [accession no. DRA009830].

AUTHOR CONTRIBUTIONS

JiS analyzed the data in consultation with KS. JiS and RS-I wrote the manuscript. RS-I, AH, KKS, and JuS refined the manuscript. HH performed the experiments. AH, KKS, and JuS supervised the project. All authors read, corrected, and approved the manuscript.

FUNDING

This study was funded by the Swiss National Science Foundation to RS-I (Marie-Heim Högtlin grant) and KKS (31003A_182318); by University Research Priority Programs, Evolution in Action of the University of Zurich to RS-I and KKS; by the Human Frontier Science Program to KKS, AH, and JuS; by the Japan Science and Technology Agency, Core Research for Evolutionary Science and Technology grant number JPMJCR16O3, Japan; and by KAKENHI grant numbers 16H06469 to JuS, KKS, and JiS.

ACKNOWLEDGMENTS

We would like to thank Walter Brücker for the support in the fieldwork in Urnerboden, Martin Lysak, Terezie Mandáková, Karol Marhold, Judita Zozomová-Lihová, Pamela Soltis, and Douglas Soltis for discussion. Data analysis was partially performed on National Institute of Genetics (NIG) at Research Organization of Information and Systems (ROIS), Japan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.567262/full#supplementary-material

Supplementary Figure 1 | Leaf vivipary in a representative leaf of *C. rivularis.* Anew plantlet was first visible 96 hours after submergence (circled, shown in close-up below). Plantlets initiated from dormant shoot meristems (shoot meristem with visible leaf shown). Shoot growth was detected 8 hours after submergence, followed by root initiation (16 hours). The shoot and root poles appeared to fuse (72 hours), followed by shoot and root growth to produce a plantlet (96 hours).

Supplementary Figure 2 | Overview of homeolog expression analysis. **(A)** To monitor homeolog expression profiles during submergence responses, we sequenced 27 RNA samples extracted from the leaf of the three species (*C. insueta*, *C. amara*, and *C. rivularis*) at the nine time points after the start of submergence. **(B)** A-genome and R-genome were assembled from RNA-Seq reads of the nine *C. amara* and *C. rivularis* samples, respectively. **(C)** Homeolog-specific expression was quantified using HomeoRoq pipeline. For each *C. insueta* sample, reads were mapped onto both A-genome and R-genome. Then, homeolog-specific reads were classified into *A-origin*, *R-origin* and unclassified (with the same mismatch rates on A and R) according to the number of mismatches on the two mapping results. After classification, the read count data, FPKM (fragments per kilobase of exon per million reads mapped), and A-origin ratio were calculated from the results.

Supplementary Figure 3 | Examples of mapping alignments. Alignments of RNA-seq read mapping of the four homeologs *ETR1*, *PIN3*, *CCA1*, and *PDF1* at 0 hr are visualized with Integrative Genomics Viewer (IGV). The top panel of each subfigure shows the alignments of *C. amara* reads and *A-origin* reads of *C. insueta* mapped to A-genome. The bottom panel shows the alignments of *C. rivularis* reads and *R-origin* reads of *C. insueta* mapped to R-genome.

Supplementary Figure 4 | Expression ratio between A- and R-homeologs during submergence treatment in the triploid *C. insueta*. Each dot shows the relation between the $\log_{10}$-transformed A-origin and R-origin reads of a homeolog pair at nine time points. Only the homeolog pairs with FPKM > 1.0 in either I_A or I_R samples are shown. The orange line represents the ratio A:R=1:2.

Supplementary Figure 5 | Distributions of A-origin ratio of *C. insueta* at nine time points. The width of the bin is 0.025 in these histograms. The vertical orange lines indicate one-third of A-origin ratios. The number of homeologs for plotting histograms is shown in the title of each histogram.

Supplementary Figure 6 | Time-course changes of homeolog expression. Expression profiles of *ERF1* and *CCA1*. Top panels represent expression of the four homeologs in the I_A, I_R, *C. amara*, and *C. rivularis* samples.

Supplementary Figure 7 | Overlaps of the number of VEH genes/homeologs of four genomes/subgenomes.

Supplementary Figure 8 | Time-course changes of homeolog expression of plantlet associated homeologs. Expression profiles of known meristem associated homeologs are shown in line charts. Red line indicates that the gene/homeolog was identified as VEH.

Supplementary Table 1 | Pearson correlation coefficients of A-origin ratios. Pearson correlation coefficients of A-origin ratios between two *C. insueta* samples among the nine time points.

Supplementary Dataset 1 | Videos for visualizing of plantlet initiation of *C. insueta* Video file (MP4) visualizes the plantlet initiation of *C. insueta* at the incubator under the standard condition with the 16 h light and 8 h dark. Leaflet were detached form *C. insueta* individuals and floated on the water in a beaker. Photos were taken between May 16, 2018 and June 07, 2018, with 90 min intervals. Photos taken at daylight were concatenated into video.

Supplementary Dataset 2 | Statistics of RNA-Seq data processing An Excel format file that contains the statistics of RNA-Seq data processing with HomeoRoq pipeline. Sheet 1: Number of read pairs before and after quality controls with Trimmomatic; Sheet 2: Number of reads that were mapped onto A-genome and R-genome, and number of reads that were classified into A-origin, R-origin, and unclassified reads with HomeoRoq.

Supplementary Dataset 3 | FPKM and CV of variably expressed homeologs an Excel format file that contains gene names, averages of log10-transformed FPKM, and coefficient of variation (CV) of log10-transformed FPKM of variably expressed homeologs (VEHs). Sheet 1: VEHs of IA samples;

Sheet 2: VEHs of IR samples; Sheet 3: VEHs of *C. amara* samples; Sheet 4: VEHs of *C. rivularis* samples.

Supplementary Dataset 4 | Enriched GO terms of variably expressed homeologs an Excel format file that contains gene ontology (GO) enrichment analysis results of VEHs. Sheet 1: Summarization of GO enrichment analysis results. Values in cells represent negative log10(*q*-value); Sheet IA: GO enrichment analysis result of VEHs of I_A samples; Sheet IR: GO enrichment analysis result of VEHs of I_R samples; Sheet A: GO enrichment analysis result of VEHs of *C. amara* samples; Sheet R: GO enrichment analysis result of VEHs of *C. rivularis* samples.

Supplementary Dataset 5 | Relative gene expression levels of the genes of the category GO:0009414 (response to water deprivation).

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26, 229–246. doi: 10.1111/j.1420-9101.2012.02599.x

Abbott, R., and Andrew, J. (2004). Origins, establishment and evolution of new polyploid species: *Senecio cambrensis* and *S. eboracensis* in the British Isles. *Biol. J. Linn. Soc.* 82, 467–474. doi: 10.1111/j.1095-8312.2004.00333.x

Abe, M., Takahashi, T., and Komeda, Y. (1999). Cloning and characterization of an L1 layer-specific gene in *Arabidopsis thaliana*. *Plant Cell Physiol.* 40, 571–580. doi: 10.1093/oxfordjournals.pcp.a029579

Ainouche, M. L., Baumel, A., and Salmon, A. (2004). Spartina anglica C. E. Hubbard: a natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biol. J. Linn. Soc.* 82, 475–484. doi: 10.1111/j.1095-8312.2004.00334.x

Akama, S., Shimizu-Inatsugi, R., Shimizu, K. K., and Sese, J. (2014). Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. *Nucleic Acids Res.* 42, 1–15. doi: 10.1093/nar/gkt1376

Akiyama, R., Sun, J., Hatakeyama, M., Lischer, H. E. L., Briskine, R. V., Hay, A., et al. (2019). Fine-scale ecological and transcriptomic data reveal niche differentiation of an allopolyploid from diploid parents in *Cardamine*. *bioRxiv* [Preprint]. doi: 10.1101/600783

Alabadí, D., Oyama, T., Yanovsky, M. J., Harmon, F. G., Más, P., and Kay, S. A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* 293, 880–883. doi: 10.1126/science.1061320

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

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bretagnolle, F., and Thompson, J. D. (1995). Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytol.* 129, 1–22. doi: 10.1111/j.1469-8137.1995.tb03005.x

Chao, Q., Rothenberg, M., Solano, R., Roman, G., Terzaghi, W., and Ecker†, J. R. (1997). Activation of the ethylene gas response pathway in *Arabidopsis* by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell* 89, 1133–1144. doi: 10.1016/S0092-8674(00)80300-1

Chuck, G., Lincoln, C., and Hake, S. (1996). *KNAT1* induces lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell* 8, 1277–1289. doi: 10.1105/tpc.8.8.1277

Cifuentes, M., Grandont, L., Moore, G., Chèvre, A. M., and Jenczewski, E. (2010). Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol.* 186, 29–36. doi: 10.1111/j.1469-8137.2009.03084.x

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

Czechowski, T., Stitt, M., Altmann, T., and Udvardi, M. K. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization. *Society* 139, 5–17. doi: 10.1104/pp.105.063743.1

Dickinson, T. A. (1978). Epiphylly in angiosperms. *Bot. Rev.* 44, 181–232. doi: 10.1007/BF02919079

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dong, S., and Adams, K. L. (2011). Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol.* 190, 1045–1057. doi: 10.1111/j.1469-8137.2011.03650.x

Douglas, M. G., Gos, G., Steige, A. K., Salcedo, A., Holm, K., Josephs, B. E., et al. (2015). Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc. Natl. Acad. Sci.* 112, 2806–2811. doi: 10.1073/pnas.1412277112

Drisch, R. C., and Stahl, Y. (2015). Function and regulation of transcription factors involved in root apical meristem and stem cell maintenance. *Front. Plant Sci.* 6:505. doi: 10.3389/fpls.2015.00505

Gan, X., Hay, A., Kwantes, M., Haberer, G., Hallab, A., Ioio, R. D., et al. (2016). The Cardamine hirsuta genome offers insight into the evolution of morphological diversity. *Nat. plants* 2:16167. doi: 10.1038/nplants.2016.167

Garcês, H. M. P., Champagne, C. E. M., Townsley, B. T., Park, S., Malhó, R., Pedroso, M. C., et al. (2007). Evolution of asexual reproduction in leaves of the genus *Kalanchoë*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15578–15583. doi: 10.1073/pnas.0704105104

Hay, A., and Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. *Development* 137, 3153–3165. doi: 10.1242/dev.030049

Howard, H. W. (1948). Chromosome number of *Cardamine pratensis*. *Nature* 161, 277–277. doi: 10.1038/161277a0

Husband, B. C. (2004). The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biol. J. Linn. Soc.* 82, 537–546. doi: 10.1111/j.1095-8312.2004.00339.x

Hussein, F. (1948). Chromosome number of *Cardamine pratensis*. *Nature* 161, 1015–1015. doi: 10.1038/1611015a0

Kuo, T. C. Y., Hatakeyama, M., Tameshige, T., Shimizu, K. K., and Sese, J. (2020). Homeolog expression quantification methods for allopolyploids. *Brief. Bioinform.* 21, 395–407. doi: 10.1093/bib/bby121

Levin, D. A. (2002). *The Role of Chromosomal Change in Plant Evolution*. Oxford: Oxford University Press.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Long, J. A., Moan, E. I., Medford, J. I., and Barton, M. K. (1996). A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* 379, 66–69. doi: 10.1038/379066a0

Mable, B. K. (2003). Breaking down taxonomic barriers in polyploidy research. *Trends Plant Sci.* 8, 582–590. doi: 10.1016/j.tplants.2003.10.006

Mandáková, T., Kovařík, A., Zozomová-Lihová, J., Shimizu-Inatsugi, R., Shimizu, K. K., Mummenhoff, K., et al. (2013). The more the merrier: recent hybridization and polyploidy in *Cardamine*. *Plant Cell* 25, 3280–3295. doi: 10.1105/tpc.113.114405

Mandáková, T., Marhold, K., and Lysak, M. A. (2014). The widespread crucifer species *Cardamine flexuosa* is an allotetraploid with a conserved subgenomic structure. *New Phytol.* 201, 982–992. doi: 10.1111/nph.12567

Mason, A. S., and Pires, J. C. (2015). Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet.* 31, 5–10. doi: 10.1016/j.tig.2014.09.011

Paape, T., Hatakeyama, M., Shimizu-Inatsugi, R., Cereghetti, T., Onda, Y., Kenta, T., et al. (2016). Conserved but attenuated parental gene expression in allopolyploids: constitutive zinc hyperaccumulation in the allotetraploid *Arabidopsis kamchatica*. *Mol. Biol. Evol.* 33, 2781–2800. doi: 10.1093/molbev/msw141

Planet, E., Attolini, C. S. O., Reina, O., Flores, O., and Rossell, D. (2012). htSeqTools: high-throughput sequencing quality control, processing and visualization in R. *Bioinformatics* 28, 589–590. doi: 10.1093/bioinformatics/btr700

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:eaar6089. doi: 10.1126/science.aar6089

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Rossmann, S., Kohlen, W., Hasson, A., and Theres, K. (2015). Lateral suppressor and Goblet act in hierarchical order to regulate ectopic meristem formation at the base of tomato leaflets. *Plant J.* 81, 837–848. doi: 10.1111/tpj.12782

Salisbury, E. J. (1965). The reproduction of *Cardamine pratensis* L. and *Cardamine palustris* Peterman particularly in relation to their specialized foliar vivipary, and its deflexion of the constraints of natural selection. *Proc. R. Soc. Lon. Ser. B. Biol. Sci.* 163, 321–342. doi: 10.1098/rspb.1965.0072

Sarkar, A. K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., et al. (2007). Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* 446, 811–814. doi: 10.1038/nature05703

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., and Shimizu, K. K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* 26, 193–207. doi: 10.1111/mec.13738

Shin, H., Shannon, C. P., Fishbane, N., Ruan, J., Zhou, M., Balshaw, R., et al. (2014). Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS One* 9:e0091041. doi: 10.1371/journal.pone.0091041

Smith, J. E. (1825). *The English Flora. III*. London: Longman, Hurst, Rees, Orme, Brown, and Green. doi: 10.5962/bhl.title.6340

Solano, R., Stepanova, A., Chao, Q., and Ecker, J. R. (1998). Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* 12, 3703–3714. doi: 10.1101/gad.12.23.3703

Soltis, D. E., Soltis, P. S., Pires, J. C., Kovarik, A., Tate, J. A., and Mavrodiev, E. (2004). Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biol. J. Linn. Soc.* 82, 485–501. doi: 10.1111/j.1095-8312.2004.00335.x

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Tayalé, A., and Parisod, C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140, 79–96. doi: 10.1159/000351318

Urbanska, K. M. (1977). Reproduction in natural triploid hybrids (2n=24) between *Cardamine rivularis* Schur and *C. amara* L. er. *Geobot. Inst. ETH Stift. Rubel* 44, 42–85.

Urbanska, K. M., Hurka, H., Landolt, E., Neuffer, B., and Mummenhoff, K. (1997). Hybridization and evolution in *Cardamine* (*Brassicaceae*) at urnerboden, central Switzerland: biosystematic and molecular evidence. *Plant Syst. Evol.* 204, 233–256. doi: 10.1007/BF00989208

Urbanska-Worytkiewicz, K., and Landolt, E. (1972). Natürliche Bastarde zwischen *Cardamine amara* L . und C . rivularis Schur aus den Schweizer Alpen. *Ber. Geobot. Inst. ETH. Stift. Rübel.* 41, 88–101. doi: 10.5169/seals-377675

Urbanska-Worytkiewicz, K., and Landolt, E. (1974a). Biosystematic investigations in Cardamine pratensis L.s.I. I. Diploid taxa from Central Europe and their fertility relationships. Berichte des Geobot. Instituts der ETH Stift. *Rübel* 42, 42–139.

Urbanska-Worytkiewicz, K., and Landolt, E. (1974b). Hybridation naturelle entre *Cardamine rivularis* SCHUR et *C. amara* L., ses aspects cytologiques et écologiques. *Act. Soc. Helv. Sci. Nat.* 1974, 89–90.

van Wesemael, J., Hueber, Y., Kissel, E., Campos, N., Swennen, R., and Carpentier, S. (2018). Homeolog expression analysis in an allotriploid non-model crop via integration of transcriptomics and proteomics. *Sci. Rep.* 8:1353. doi: 10.1038/s41598-018-19684-5

Vollbrecht, E., Veit, B., Sinha, N., and Hake, S. (1991). The developmental gene *Knotted*-1 is a member of a maize homeobox gene family. *Nature* 350, 241–243. doi: 10.1038/350241a0

Williams, R. W. (1998). Plant homeobox genes: many functions stem from a common motif. *BioEssays* 20, 280–282. doi: 10.1002/(SICI)1521-1878(199804)20:4<280::AID-BIES2>3.0.CO;2-U

Yamaguchi, R., Imoto, S., Yamauchi, M., Nagasaki, M., Yoshida, R., Shimamura, T., et al. (2008). Predicting differences in gene regulatory systems by state space models. *Genome Inform.* 21, 101–113. doi: 10.11234/gi1990.21.101

Yang, J., Liu, D., Wang, X., Ji, C., and Cheng, F. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* 48, 1225–1232. doi: 10.1038/ng.3657

Yoo, M.-J., Szadkowski, E., and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhao, J., Yang, F., Feng, J., Wang, Y., Lachenbruch, B., Wang, J., et al. (2017). Genome-wide constitutively expressed gene analysis and new reference gene selection based on transcriptome data: a case study from poplar/canker disease interaction. *Front. Plant Sci.* 8:1876. doi: 10.3389/fpls.2017.01876

Zielinski, M.-L., and Mittelsten Scheid, O. (2012). "Meiosis in Polyploid Plants," in *Polyploidy and Genome Evolution*, eds P. S. Soltis, and D. E. Soltis (Berlin: Springer Berlin Heidelberg), doi: 10.1007/978-3-642-31442-1

Zozomová-Lihová, J., Krak, K., Mandáková, T., Shimizu, K. K., Španiel, S., Vít, P., et al. (2014). Multiple hybridization events in *Cardamine* (*Brassicaceae*) during the last 150 years: revisiting a textbook example of neoallopolyploidy. *Ann. Bot.* 113, 817–830. doi: 10.1093/aob/mcu012

Conflict of Interest: JuS is the CEO of the company Humanome Lab.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 November 2020
doi: 10.3389/fgene.2020.596150

Expression Partitioning of Duplicate Genes at Single Cell Resolution in *Arabidopsis* Roots

Jeremy E. Coate[1†‡], Andrew D. Farmer[2†‡], John W. Schiefelbein[3†] and Jeff J. Doyle[4†]*

[1] Department of Biology, Reed College, Portland, OR, United States, [2] National Center for Genome Resources, Santa Fe, NM, United States, [3] Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, United States, [4] School of Integrative Plant Science, Plant Biology Section, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:
Jonathan F. Wendel,
Iowa State University, United States

Reviewed by:
Kentaro K. Shimizu,
University of Zurich, Switzerland
Nicholas Louis Panchy,
The University of Tennessee,
Knoxville, United States

***Correspondence:**
Jeff J. Doyle
jjd5@cornell.edu

†ORCID:
Jeremy Coate
orcid.org/0000-0002-1619-8643
Andrew Farmer
orcid.org/0000-0002-4224-2433
John Schiefelbein
orcid.org/0000-0002-0560-5872
Jeff Doyle
orcid.org/0000-0003-1579-9380

‡These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 18 August 2020
Accepted: 12 October 2020
Published: 03 November 2020

Citation:
Coate JE, Farmer AD, Schiefelbein JW and Doyle JJ (2020) Expression Partitioning of Duplicate Genes at Single Cell Resolution in Arabidopsis Roots. Front. Genet. 11:596150. doi: 10.3389/fgene.2020.596150

Gene duplication is a key evolutionary phenomenon, prevalent in all organisms but particularly so in plants, where whole genome duplication (WGD; polyploidy) is a major force in genome evolution. Much effort has been expended in attempting to understand the evolution of duplicate genes, addressing such questions as why some paralog pairs rapidly return to single copy status whereas, in other pairs, both paralogs are retained and may diverge in expression pattern or function. The effect of a gene – its site of expression and thus the initial locus of its function – occurs at the level of a cell comprising a single cell type at a given state of the cell's development. Using *Arabidopsis thaliana* single cell transcriptomic data we categorized patterns of expression for 11,470 duplicate gene pairs across 36 cell clusters comprising nine cell types and their developmental states. Among these 11,470 pairs, 10,187 (88.8%) had at least one copy expressed in at least one of the 36 cell clusters. Pairs produced by WGD more often had both paralogs expressed in root cells than did pairs produced by small scale duplications. Three quarters of gene pairs expressed in the 36 cell clusters (7,608/10,187) showed extreme expression bias in at least one cluster, including 352 cases of reciprocal bias, a pattern consistent with expression subfunctionalization. More than twice as many pairs showed reciprocal expression bias between cell states than between cell types or between roots and leaves. A group of 33 gene pairs with reciprocal expression bias showed evidence of concerted divergence of gene networks in stele vs. epidermis. Pairs with both paralogs expressed without bias were less likely to have paralogs with divergent mutant phenotypes; such bias-free pairs showed evidence of preservation by maintenance of dosage balance. Overall, we found considerable evidence of shifts in gene expression following duplication, including in >80% of pairs encoding 7,653 genes expressed ubiquitously in all root cell types and states for which we inferred the polarity of change.

Keywords: gene duplication, single cell RNA-seq, cell type, cell state, polyploidy, expression subfunctionalization

INTRODUCTION

According to Lynch and Trickovic (2020, p. 1861), "One of the last uncharted territories in evolutionary biology concerns the link with cell biology. Because all phenotypes ultimately derive from events at the cellular level, this connection is essential to building a mechanism-based theory of evolution." As a candidate for building such a connection to cell biology, it would be difficult to

identify a more important molecular evolutionary process than gene duplication, whose key role has been universally recognized since the classic paper of Ohno (1970), half a century ago. Gene duplication occurs at high frequency, estimated at 0.01 duplications per gene per million years in eukaryotic genomes, and large numbers of recently formed paralogs are found in typical animal, fungal, and plant genomes (Lynch and Conery, 2000; Lynch et al., 2001). Among eukaryotes, plant genomes, in particular, are characterized by massive levels of duplication, thanks to waves of whole genome duplication (WGD, polyploidy). Recent estimates place the number of known plant WGD events at the genus level or above at over 250, and most plant lineages have experienced multiple cycles of polyploidy (Van de Peer et al., 2017; Leebens-Mack et al., 2019). Because they are the products of both small scale duplications (SSD) and WGD, plant gene families can be very large and complex (Panchy et al., 2016).

The fate of most paralogs, whether produced by SSD or WGD, is pseudogenization and eventual loss, through mutations that inactivate redundant copies during the "fixation phase" of a duplicate gene's life cycle (Innan and Kondrashov, 2010; Xie et al., 2019). Various mechanisms have been hypothesized that can preserve paralog pairs by making both copies of the gene indispensable (Innan and Kondrashov, 2010; Panchy et al., 2016; Qiao et al., 2019). These mechanisms can differ for SSD vs. WGD, even operating in different directions in the case of dosage effects (Papp et al., 2003; Freeling, 2009; Birchler and Veitia, 2010, 2012, 2014). Understanding why and how gene pairs are retained is complicated in part because in many cases competing hypotheses are difficult to distinguish from one another in terms of their predictions (Innan and Kondrashov, 2010). Obtaining empirical data for testing these hypotheses is not easy. Several of the models involve "function," a term that can be difficult to define.

Gene expression is often used as a proxy for gene function when assessing the fates of duplicate genes (e.g., Panchy et al., 2019), with biased expression of paralogs of a duplicated gene providing evidence for sub- or neofunctionalization. Expression proportional to gene copy number is a key component of models that involve preservation of duplicates via stoichiometric constraints (Coate et al., 2016; Song et al., 2020). Expression occurs in the nuclei of individual cells, which comprise different "cell types," the existence of which is taken as a given in the molecular and developmental biology literature, but which are difficult to define. Although there has been recent theoretical progress in how cell types originate and evolve (e.g., Arendt et al., 2016; Liang et al., 2018; Yuan et al., 2020), Vickaryous and Hall (2006) liken the problem of defining "cell type" to defining "species" – an endless source of controversy in evolutionary biology. One major complication is that although cell identity may be stable, "the same cell type can exhibit a range of different phenotypes (states)" (Morris, 2019) in response to diverse physiological or developmental stimuli. Each state of each cell type has its own characteristic transcriptome, so it is at the level of individual cell types and states that we should find the most precise transcriptomic data to explore the expression-based mechanisms that preserve duplicate genes. Studies at the tissue or organ level aggregate different cell types, obscuring patterns of gene expression that may be of interest (**Figure 1**; Efroni and Birnbaum, 2016; Libault et al., 2017).

Until recently, plant studies at the single cell type level were mostly limited to isolation of individual cell types by flow sorting (e.g., Birnbaum et al., 2003), or to cell types for which large populations of pure cells can be obtained, such as root hairs (e.g., Qiao and Libault, 2013; Hossain et al., 2015) and cotton seed fibers (e.g., Shi et al., 2006; Gou et al., 2007; Taliercio and Boykin, 2007). Cotton fibers have been studied extensively in the context of polyploidy (e.g., Hovav et al., 2008a; Yoo and Wendel, 2014; Gallagher et al., 2020). However, bulked samples of a single cell type, even when collected at different developmental stages, may still miss some details of transitions among cell states. Additionally, by focusing on a single cell type, such studies are blind to any expression partitioning between paralogs that might have occurred across cell types.

Single cell methods that have revolutionized biology continue to develop and promise ever more powerful and precise data (Lähnemann et al., 2020). In plants, several groups recently published single cell transcriptomic studies of *Arabidopsis* roots (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019) that not only identified known cell types, including cell types represented by small numbers of cells that would be missed in conventional transcriptomic studies, but also revealed cells with distinctive transcriptomes not readily assigned to known cell types, and subdivided cell types into different developmental states. These data provide a potential resource for exploring gene duplication events at the single cell level in *A. thaliana*. Much is known about gene duplication in this model species, including classification of the origins of its thousands of paralogous gene pairs by various mechanisms of single gene duplication (SSD) and whole genome duplication (WGD, polyploidy) (e.g., Wang et al., 2013; Hao et al., 2018; Qiao et al., 2019) and the degree to which paralogs from many pairs have diverged functionally (Hanada et al., 2009; Panchy et al., 2019). Much is also understood about the process of biased genome fractionation following WGD 30–40 million years ago (MYA), which has led to the retention of only a subset of duplicated genes, unequally distributed across the homoeologous subgenomes of *A. thaliana* (e.g., Cheng et al., 2018; Emery et al., 2018; Liang and Schnable, 2018).

The availability of *Arabidopsis* root single cell data allowed us to explore expression patterns of over 11,000 paralogous gene pairs at a finer scale than has previously been reported. We find many examples of expression differentiation of paralogous genes at the level of cell types and states within a single organ, similar to what Adams et al. (2003) found at the level of whole organs comprising the flower. A large fraction of gene pairs show evidence of evolutionary shifts in expression between paralogs, including over 75% of the over 1,500 gene pairs with one or both paralogs ubiquitously expressed in all root cell types and states. Patterns of expression of gene pairs from whole genome duplications vs. single gene duplications are mostly consistent with expectations, with evolutionary differentiation of expression between paralogs more common following single

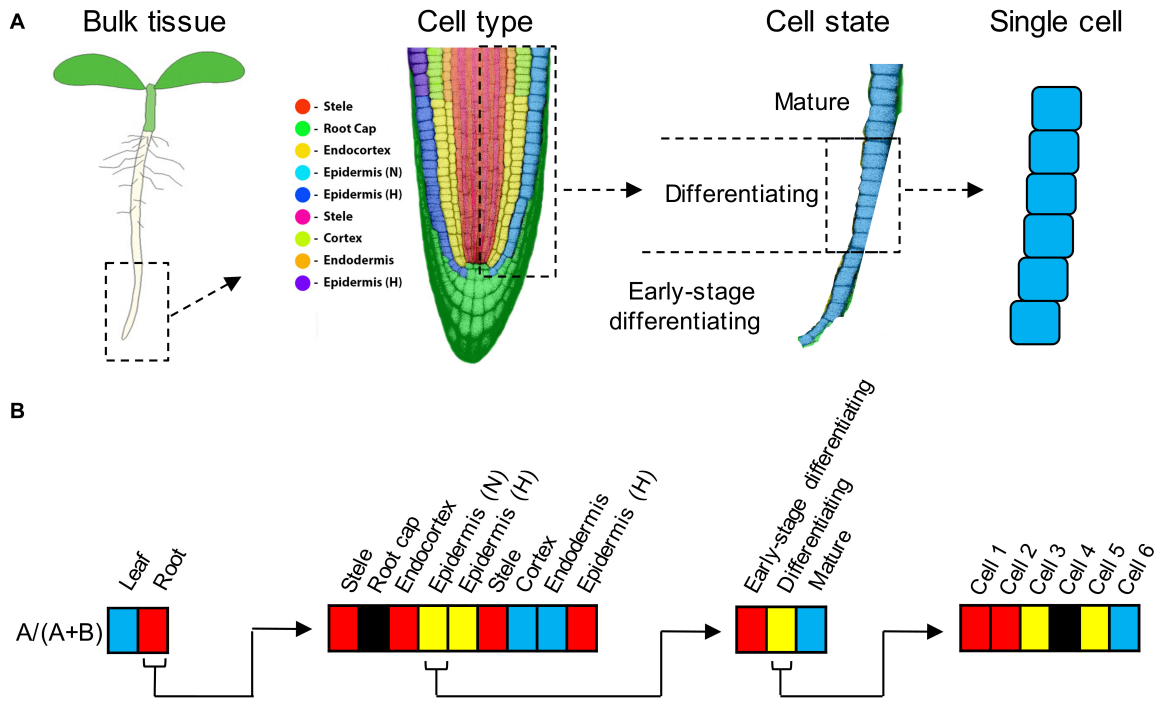


FIGURE 1 | (A) Expression subfunctionalization is the partitioning of ancestral expression profiles between paralogs. Partitioning can occur at the level of tissue types (e.g., leaves vs. roots), but also at increasingly finer scales, including at the level of cell type (in roots, roughly corresponding to the nine cell clusters identified by Ryu et al., 2019), or cell state (e.g., developmental gradients within a single cell type such as non-hair epidermal cells), and even among single cells. Under cell type, "Epidermis (N)" refers to non-hair (atrichoblast) epidermal cells, and "Epidermis (H)" refers to root hair (trichoblast) epidermal cells (Ryu et al., 2019). There are two clusters labeled "Epidermis (H)" because the available marker genes assigned them both to the same cell type, but their transcriptional profiles were sufficiently distinct (most likely due to differences in developmental stage) that the clustering algorithm separated them. There are two clusters labeled "Stele" for similar reasons. **(B)** Examples of how expression partitioning between paralogs (designated "A" and "B") can occur at different levels. Expression bias is indicated by shading (red = expression biased toward paralog A [specifically, A/(A + B) > 0.9], blue = expression biased toward paralog B [A/(A + B) < 0.1], yellow = unbiased expression, black = neither paralog expressed). Left, in this example, paralog A expression predominates in roots, whereas paralog B expression predominates in leaves. Middle left, within root tissue, paralogs A and B exhibit partitioning by cell type. Middle right, within the non-hair epidermal cell type, paralogs A and B exhibit partitioning by cell state (i.e., developmental stage). Right, within the "differentiating" stage of non-hair epidermal cell type, A and B paralog expression is partitioned among individual cells (paralog A predominating in cells 1–2, and paralog B expression predominating in cell 6).

gene duplications than following polyploidy, and pairs from WGD events showing evidence of preservation by dosage balance (e.g., Freeling, 2009; Panchy et al., 2016; Tasdighian et al., 2017; Defoort et al., 2019; Qiao et al., 2019). We find evidence of concerted divergence of gene networks between different root cell types. We also show that different cell types have responded differently to gene and genome duplications in the degree to which they deploy one or both paralogs in their transcriptomes.

MATERIALS AND METHODS

Single Cell Datasets

Illumina sequence data for NCBI SRA experiments SRX5074330-SRX5074332 corresponding to scRNA-seq data from three wild-type replicates of *Arabidopsis* roots from Ryu et al. (2019) were processed with the 10× Genomics Cell Ranger v3.1.0 count pipeline, run independently on the data for each replicate to produce unique molecular identifier (UMI) raw counts matrices against the TAIR10 genome using Araport11 annotations (Cheng et al., 2017). Custom scripts (available from[1]) were used to produce per-cluster UMI counts for each gene, summing the contributions from all cells assigned to the 9 "superclusters" presented in Ryu et al. (2019) and separately for 36 root cell clusters ("RCCs") derived from those 9 initial superclusters using the Seurat software package (Butler et al., 2018; FindClusters function) with default parameters (perplexity = 30, random seed = 1) and a resolution of 3.5. These per-supercluster and per-RCC gene UMI count matrices formed the basis of subsequent analysis of expression bias between duplicated gene pairs. Specific cell types and differentiation states were assigned to each of the 36 RCCs from the 9 initial superclusters (**Supplementary Table 1**) using previously defined marker genes (Ryu et al., 2019).

To filter out spurious expression signals (resulting, for example, from doublets or from cell-free RNA), we required at least one UMI from two or more cells in a given cluster for a gene to be considered expressed in the context represented by the cluster. In some cases, we also analyzed the data using the

[1]https://github.com/adf-ncgr/singlecell_paralogue_expression_scripts

minimally restrictive expression threshold of ≥1 UMI from ≥1 cell to assess how strongly additional filtering affects the results.

The cells in the Ryu et al. (2019) study from which the data were taken were derived from protoplasted root tissues, and would thus be subject to some level of protoplasting-induced changes in gene expression relative to untreated tissues (Birnbaum et al., 2003). We considered that altered responses to the protoplasting treatment were within the scope of what could be considered paralog expression divergence, and we retained duplicate pairs involving such genes in the subsequent analyses.

Sequence Read Archive (SRA) Datasets

A total of 214 RNA-seq datasets for *Arabidopsis thaliana* used in Panchy et al. (2019) were obtained from NCBI SRA (Leinonen et al., 2011) and aligned to the TAIR 10 genome using hisat2 v2.1.0 (Kim et al., 2019) against an index built with splice sites and exons derived from Araport11 annotations. TPM values derived by stringtie v2.0.6 (Kovaka et al., 2019) were then converted to raw counts for the Araport 11 genes (**Supplementary Table 2**). These data were used to assess if genes not expressed in the Ryu et al. (2019) scRNA-seq data were expressed in other tissues and/or conditions. Of the 214 SRA libraries, 37 were generated from leaf tissue only (832 million reads total) and 31 were from root tissue only (290 million reads total). Counts were summed across all samples in each class (leaf vs. root) to provide high-coverage bulked data sets to compare paralog expression at the level of contrasting tissue types, in the same manner as for the clustered single-cell data, as described below.

Biased Expression of Paralogs

For gene pairs identified by Wang et al. (2013), we determined whether the paralogs showed biased expression (analogous to "biased homoeolog expression" of Grover et al., 2012) in the context of each cluster, using a UMI cutoff of 9:1. We chose 9:1 as a stringent threshold for biased expression because our primary interest was in identifying cases of extreme imbalance in paralog expression, consistent with expression subfunctionalization. The two paralogs of the gene pair were then designated "single-cell A" (scA) and "single-cell B" (scB), with scA being the dominant paralog showing higher expression in the greatest number of clusters.

In order to exclude from consideration gene pairs whose apparent bias may be insignificant relative to random sampling deviations, we considered the counts characterizing each gene pair as representing the outcome of a Bernoulli trial, where the abundance of reads relative to its partner determines the probability of each outcome (i.e., sampling a particular partner when a read is chosen from one of the pair). In order to determine confidence in the estimate of the probability given a specific number of trials (i.e., the summed read count for the pair), the Wilson (1927) score interval estimation was used to provide a 95% confidence interval around the expression bias value estimated from the read counts. Considered from the perspective of the dominant gene in a putatively biased pair, the lower bound on the confidence interval can be used as the minimum level of bias at the chosen confidence level, and if this minimum level of bias falls below the 9:1 threshold ratio for considering a gene pair to exhibit bias, it was removed from consideration when calculating the fixation and balance indices described below.

For each paralog pair we calculated two indices to describe its pattern of expression across cell clusters:

- The Expression Fixation index (F_{ex}) measures the degree of bias in the expression of paralogs of a given pair across the cell clusters. $F_{ex} = N_{fix}$/(number of cell clusters expressing at least one paralog above a cutoff threshold), where N_{fix} is the number of clusters for which one paralog (either one) is "fixed" (is preferentially expressed at or above the 9:1 threshold).
- The Balance Index (B_{fix}) is calculated for any paralog pair for which at least one cluster is fixed for a paralog, and measures the degree to which one paralog dominates the expression across the clusters. $B_{fix} = 2^*$(number of clusters fixed for scB)/(number of clusters fixed for either paralog). B_{fix} runs from 0–1; cases with no fixation of either paralog (paralog pairs with $F_{ex} = 0$) have no B_{fix} score (N/A).

Examples are given in **Figure 2**, with explicit calculations of the indices given for the gene pairs in panel B and the continuous color scale at the right indicating the numerical values (black indicates an undefined value due to a zero appearing in the denominator). As may be seen by comparing panels A and B of **Figure 2**, which represent several gene pairs characterized at both the 36 RCC level (**Figure 2A**) and the 9 supercluster level (**Figure 2B**), the values of the indices can differ relative to the granularity of the clustering at which a given gene pair is considered. In general, the finer resolution of the 36 RCC allows more sensitivity in observing biased contexts that would otherwise be obscured by aggregation with other less-biased or oppositely-biased contexts (e.g., the last row in each table representing gene pair AT3G18950/AT1G49450). In contrast, in some cases going to a finer level of clustering can actually diminish the statistical power of detecting bias due to the lower UMI counts associated with each of the individual RCCs (e.g., the fifth row in each table representing gene pair AT3G18350/AT1G48840).

Classification of Paralogs by Expression Pattern

We assigned paralog pairs to one of five classes based on their expression patterns across the 36 RCCs (**Table 1** and **Figure 2A**). Genes were also assigned to equivalent classes at the nine cluster level (e.g., Class 1 if only one of the two copies was expressed at the 9-supercluster level; **Figure 2B**). As noted in **Table 1**, these classes depend not only on our previously stated definition of expression/non-expression of each gene (UMI from two or more cells) but also on the values of the F_{ex} and B_{fix} indices, which characterize patterns of *statistically significant* bias. Thus, for example, in **Figure 2A**, for the Class 1 gene pairs which by definition involve expression of only the scA paralog, we see a mixture of contexts in which some are black (implying no expression of either paralog), some are red (implying biased expression meeting the Wilson criterion) and some are yellow (implying that expression did not meet the statistical test for bias,

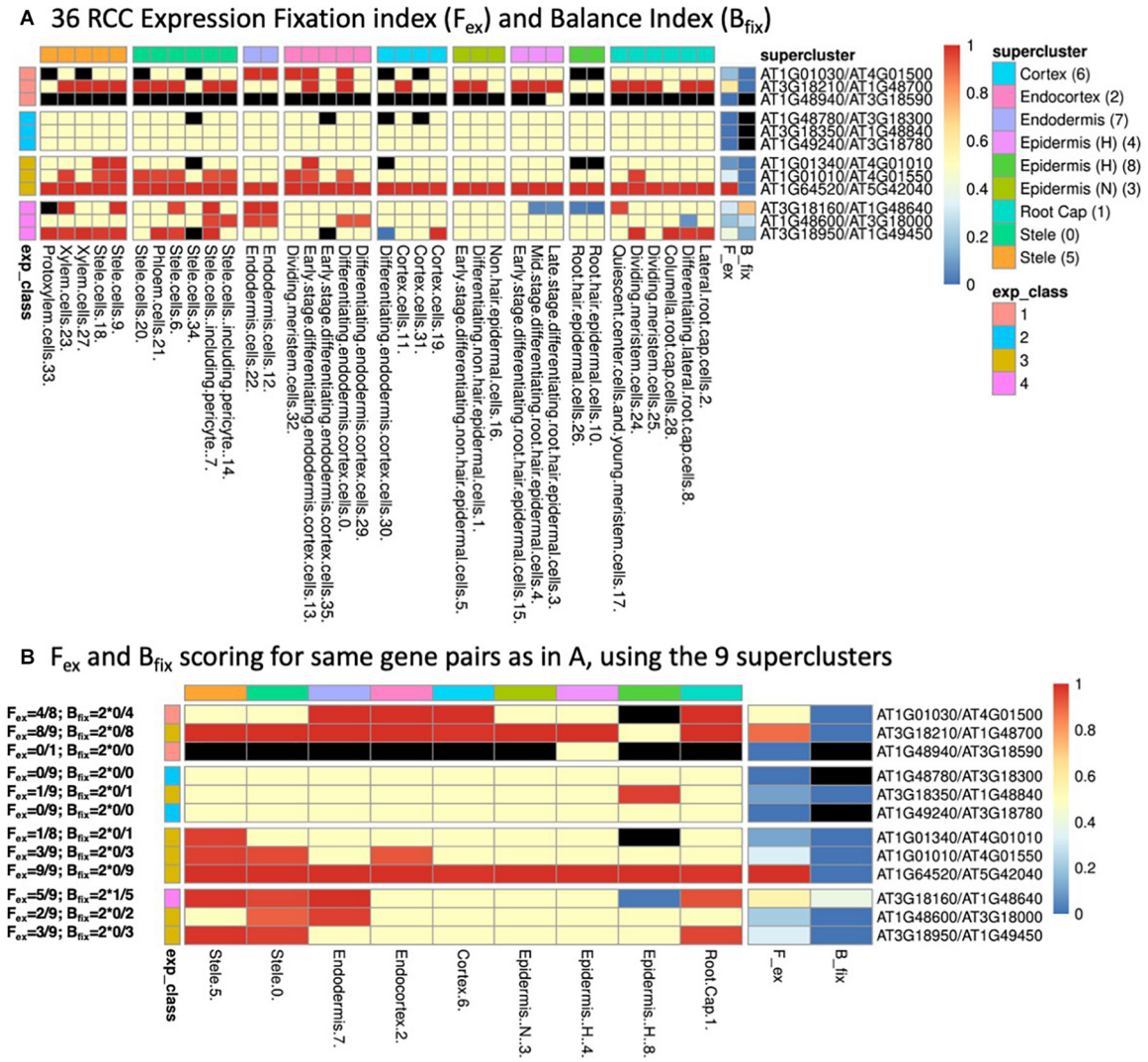


FIGURE 2 | (A) Heatmap of 36 root cell clusters (RCCs) from nine Ryu et al. (2019) superclusters, showing examples of selected gene pairs illustrating F_{ex} and B_{fix} (last two columns before gene pair designations) and the four expression classes (defined in **Table 1**). Cell colors: for RCCs, red cells have dominant paralog scA fixed (expressed above the 9:1 threshold at 95% confidence level, or, in the case of Class 1, expressed at 1:0); blue cells have paralog scB fixed; yellow cells have at least one of the paralogs expressed, but do not meet the statistical criterion for biased expression; black cells have no expression of either paralog. For F_{ex} and B_{fix} the cell values for each index run from 0–1 and are colored continuously from blue to red as indicated by the scale at the right of the figure. RCCs are grouped according to the superclusters from which they were derived (as indicated by column heading colors), while gene pairs are grouped according to their expression classes (as indicated by row heading colors). **(B)** Corresponding heatmap of the same gene pairs as shown in **(A)** using the nine Ryu et al. (2019) superclusters as the basis for assessment of expression bias. Expression class colors are as in **(A)**. Examples of the calculation of F_{ex} and B_{fix} values for each gene pair based on the expression-biased contexts detected in each case are given to the left of the figure.

despite the fact that, in the context of Class 1, only scA met the definition of expression). Again, a comparison of **Figure 2** panels A and B shows that the expression class assigned to a given gene pair can change when the level of resolution of the clusters is altered.

Fixation Similarities and Differences Across Clusters

Heatmaps were produced on the results of calculating biased expression for each gene pair across all clusters, by using the R package "Pretty Heatmaps" (pheatmap v1.0.12). The annotations option of pheatmap was used to denote the expression class (**Table 1**).

Single Cell Measurements of Paralog Usage

In order to look at possible bias between duplicated genes at the level of single cells, we again applied the Wilson (1927) test on the UMI counts for the duplicated genes at the level

TABLE 1 | Expression classes of duplicate gene pairs.

Class	Copies expressed	F_ex	B_fix	Summary
0	0	0	0	No expression in any RCC
1	1	≥0*	0	Only 1 copy expressed
2	2	0	0	Both copies expressed, no fixation
3	2	>0	0	Both copies expressed, some fixation, no balance
4	2	>0	>0	Both copies expressed, balanced fixation (subfunctionalization?)

**In some Class 1 cases, despite only one copy being expressed, no clusters have significant fixation due to low read counts.*

of individual cells. In this case, due to the low counts obtained and in order not to confute subsequent tests we extended the states recorded by the test to distinguish not only between biased and unbiased but also cases where the counts were too low to determine whether a particular gene pair fell in one class or another in a given cell. Cells in the latter case were excluded from subsequent analyses. For the remainder, gene pairs could be tallied with respect to the number of cells in a given cluster that showed bias for one or the other of the genes, were unbiased with respect to the expression of the genes, or showed no expression of either gene (i.e., had a UMI count of zero for both). The number of cells falling into each category were then used to test for significant overlap of the lists of cells expressing each of the two genes, using Fisher's exact test in the manner of the GeneOverlap Bioconductor package (Shen and Sinai, 2019), while additionally testing for significant non-overlap (i.e., a tendency for cells expressing one paralog to not express the other paralog) by utilizing the alternate tail of the hypergeometric distribution.

K_a/K_s Value Determination for Paralogs

In order to assess whether different classes of paralogs based on patterns of expression showed significant differences in terms of protein coding divergence, we used the implementation of the Nei-Gojobori algorithm (Nei and Gojobori, 1986) provided by the Bio:Align:DNAStatistics module of BioPerl (Stajich et al., 2002) through wrapper scripts available in the MCScanX software distribution (Wang et al., 2012) but run across all the *Arabidopsis* gene duplicate pairs classified by Wang et al. (2013) including non-syntenic SSD pairs.

Gene Ontology Enrichment Analysis

GO term enrichment was assessed using ThaleMine's GO enrichment analysis widget with default parameters. Specifically, GO term representation in specific gene sets was compared to representation in the full *Arabidopsis* gene set (Araport 11; Cheng et al., 2017), and tested for significance at $p < 0.05$ (Fisher's Exact Test with Holm-Bonferroni correction).

RESULTS

Arabidopsis Root Cell Clusters Each Express Over 35% of Genes in the Genome

We studied both the 9 root cell supercluster transcriptome data of Ryu et al. (2019) and transcriptomes of the 36 root cell clusters (RCCs) derived from those superclusters. These two datasets are strongly nested, with the 9 superclusters broadly corresponding to cell types and the 36 RCC data including developmental states of those types (**Figure 3** and **Table 1**). We primarily report results from the 36 RCCs here (9 supercluster counts data and 36 RCC counts data are available in **Supplementary Table 3**).

We explored two different thresholds for determining whether a gene was expressed in an RCC: at least 1 UMI per cluster vs. the more stringent cutoff of expression in two or more cells of a given RCC. These thresholds were chosen to accommodate different technical issues with single-cell RNA-seq droplet-based methods (Luecken and Theis, 2019). On the one hand, technical dropout (Bhargava et al., 2014) leads to reduced capture for low abundance transcripts, suggesting a relaxed threshold for counts relative to bulk RNA-seq; conversely, the possibility of capturing multiple cells of differing types in a single droplet ("doublets") suggests a need to guard against false positives generated by this phenomenon. We found that counts derived from these two thresholds were strongly correlated (R^2 = 0.997) but 7–18% lower for the two cell cutoff than for the 1 UMI cutoff (**Supplementary Figure 1**); we report numbers using the more stringent cutoff throughout.

After excluding loci annotated as "novel transcribed regions," "pseudogenes," "transposable element genes," and organelle-encoded genes, there are 32,548 genes in the most recent annotation of the *Arabidopsis* genome (Araport11: Cheng et al., 2017). Of these, 22,669 (70%) were expressed in at least one of the 36 RCCs, with each RCC expressing 35–58% of these genes (**Figure 4**). These percentages are comparable to *Arabidopsis* pollen cell stages (32–51% of microarray features; Honys and Twell, 2004), and are somewhat lower than those for cotton fiber cells, which transcribe from 75–94% of the genome's genes depending on developmental stage (Hovav et al., 2008b). Differences in the number of genes expressed per cell cluster were in many cases statistically significant, but these differences were largely driven by differences in cell count per RCC (**Supplementary Figure 2**). Approximately 35% of root-expressed genes (∼22% of all *Arabidopsis* genes) were expressed in all 36 RCCs ("RCC-ubiquitous," subsequently referred to as "RCC-u"); 1,059 genes (4.7% of root-expressed genes) were uniquely expressed in only one RCC (**Figure 4**).

Many Gene Pairs Show Biased Paralog Expression in Root Cell Clusters, and Different Duplication Types Show Different Expression Patterns

Wang et al. (2013) identified 11,630 gene pairs in the *Arabidopsis* genome, and classified them as being duplicated either by WGD

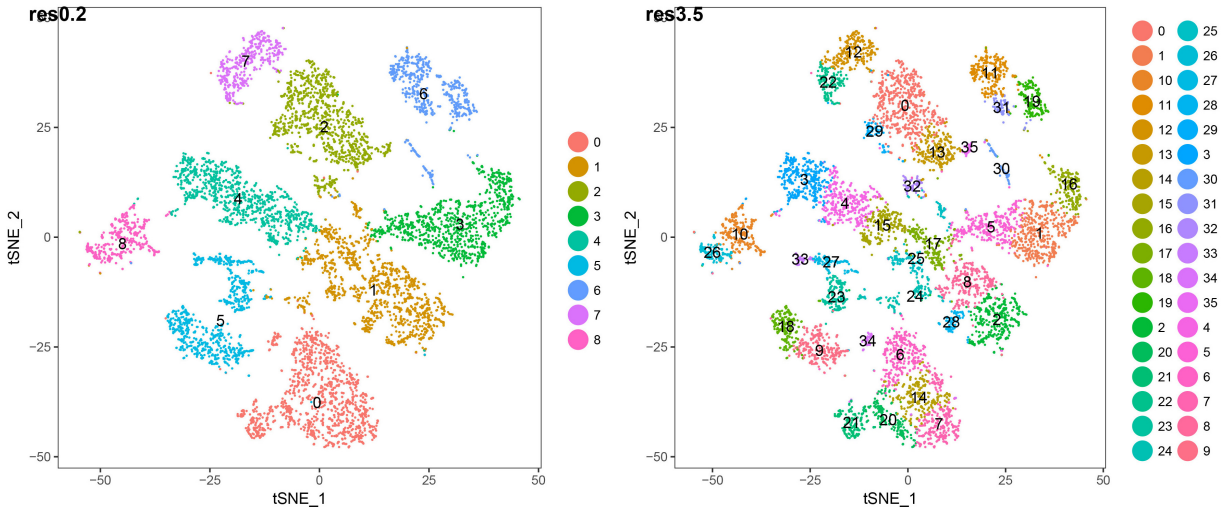


FIGURE 3 | t-SNE plots of 7522 wild-type *Arabidopsis* root single-cell transcriptomes clustered at two different resolutions. The lower resolution generates 9 major clusters ("superclusters"; roughly representing "cell types") and the higher resolution generates 36 root cell clusters (RCCs; roughly representing "cell states"). The specific cell type/state assignments for each cluster are provided in **Figure 2**. Additional details are provided in **Supplementary Table 1**.

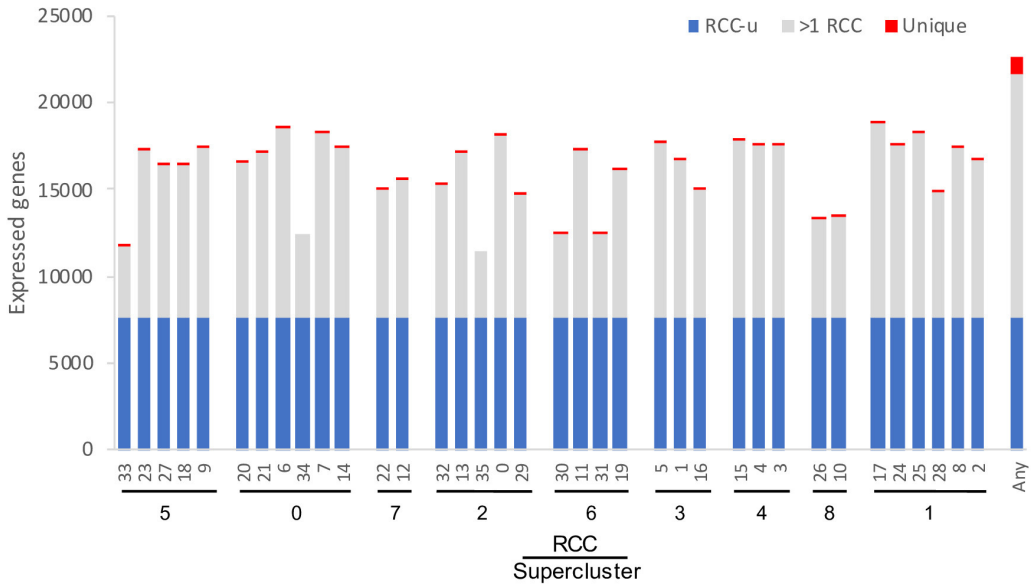


FIGURE 4 | Number of genes expressed per cluster. The 36 RCCs are grouped by Ryu et al. (2019) supercluster (roughly representing cell types and developmental states; **Figure 1**), and ordered roughly from the interior (stele) to the exterior (epidermis) and tip (root cap) of the root. RCC-ubiquitous (RCC-u) is the subset of genes expressed in all 36 RCCs, ">1 RCC" is the subset expressed in 2–35 RCCs, and "Unique" is the subset expressed in only one RCC. "Any" is the union of all genes expressed in at least one of the 36 clusters.

or by SSD (**Supplementary Table 3**). They further subdivided the WGD class into products of the alpha (most recent, around 31.8–42.8 MYA; Edger et al., 2018), beta (85–92.2 MYA; Edger et al., 2018), and gamma events (115–120 MYA; Jiao et al., 2012), and the SSD class into tandem duplicates, proximal duplicates, and two subclasses of transposed duplicates (younger than 16 MYA vs. older). 160 of the Wang et al. (2013) pairs contain obsolete gene models and were excluded from subsequent analyses. We divided the remaining 11,470 paralog pairs into five RCC expression classes according to whether neither paralog (Class 0), only one paralog (Class 1), or both paralogs (Classes 2, 3, and 4) were expressed in at least one cluster; Classes 2, 3, and 4 were distinguished from one another by patterns of biased paralog expression estimated at a 9:1 ratio (**Figure 2** and **Supplementary Table 3**).

Overall, 10,187 of these gene pairs (88.8% of total pairs) belonged to Classes 1–4, having one or both paralogs expressed in at least one RCC (**Supplementary Table 4**). This percentage

is similar to but significantly lower than the expectation for drawing at least one member of a gene pair from the 70% of *Arabidopsis* genes expressed in root cell clusters (90.8%; $\chi^2 = 24.3$, $p < 0.001$). Notably, however, 67.7% (7,764 pairs) had both copies expressed in root cell clusters, significantly higher than the random expectation of 48.5% ($\chi^2 = 867.4$, $p < 0.001$). There was a clear distinction between WGD and local SSD (proximal and tandem) pairs with regard to these percentages. WGD pairs were significantly more likely to express at least one copy ($\geq$96.8%; $\chi^2 \geq 16.23$, $p < 0.001$) and to express both copies ($\geq$79.0%; $\chi^2 \geq 101.6$, $p < 0.001$) than expected by chance, whereas local SSD pairs were significantly less likely to express at least one copy ($\leq$72.9%; $\chi^2 \geq 97.0$, $p < 0.001$) and less than or similarly likely to express both copies (proximal: 42.4%; $\chi^2 = 5.76$, $p = 0.016$; tandem: 47.9%; $\chi^2 = 0.136$, $p = 0.712$). Older duplicates created by transposable elements (TEs) exhibited a similar pattern to WGD duplicates and younger TEs exhibited a similar pattern to local SSD duplicates (**Figure 5** and **Supplementary Table 4**).

The expression breadth of pairs (number of clusters in which a pair is expressed) also differed by duplication type (**Supplementary Figure 3**). For all three WGD types, around 44–48% of paralog pairs had one or both paralogs expressed in all 36 RCCs ("RCC-u pairs"), and 30–38% of genes were RCC-u. Only 4–6% of pairs had no expression of either paralog in any RCC (Class 0 pairs; **Table 1**, **Figure 5**, and **Supplementary Figure 3**), comprising 8–12% of all genes. In contrast, tandem and proximal SSD types had no expression of 25–30% of pairs and around 35–40% of their individual genes in any RCC and fewer than 20% of pairs and 12% of their genes were expressed in all 36 RCCs. Transposed duplicates were intermediate between these two groups, with older pairs again behaving more like WGD types (44% of pairs and 30% of individual genes expressed in all RCCs; 2% of pairs and 12% of genes not expressed in any RCC) and younger transposed pairs behaving more like the other SSD types (28% of all pairs and 17% of genes expressed in all 36 RCCs; 18% of pairs and 33% of genes not expressed in any RCC).

Class 1 gene pairs have the same paralog exclusively expressed in all root cell clusters in which either paralog is expressed (balance index [B_{fix}] = 0) (**Figure 2** and **Table 1**). The same dichotomy between duplication types observed for Class 0 gene pairs was also observed for Class 1: Only around 10–20% of WGD and older transposed duplicate pairs belonged to this class, vs. 25–35% of tandem, proximal, and younger transposed SSD classes (**Figure 5** and **Supplementary Table 4**).

In a Class 1 pair, only one paralog (by definition the dominant paralog, scA) is expressed in roots. ScB shows no expression in any RCC, and could either be a pseudogene, or could be a functional gene expressed in other contexts than the roots studied here. We thus looked at expression of Class 1 scB paralogs in 214 RNA-seq experiments obtained from the SRA database (Leinonen et al., 2011; **Supplementary Table 2**). We found that in 93% of Class 1 gene pairs, the scB paralog was expressed in at least one SRA dataset, and 91% were expressed in non-root SRA libraries (requiring at least 5 reads to be considered expressed) (**Supplementary Table 5**). These results indicate that the majority of Class 1 genes (90–93%) represent cases where expression has been partitioned between paralogs since their divergence, with scA being the only copy now expressed in roots grown under the conditions used by Ryu et al. (2019). Although this suggests that many of these scB paralogs could be functional, many could still be pseudogenes, because around a third of plant pseudogenes are expressed (Xie et al., 2019). Lloyd et al. (2018) found that expression level and, particularly, expression breadth across tissues, were the best predictors of functional genes. Both of these metrics were lower on average for the Class 1 scB copies than for Class 1 scA or for all other genes in the *Arabidopsis* genome. This was true whether looking at all Class 1 pairs combined or separated by duplication mechanism (**Supplementary Figure 4**), suggesting that Class 1 scB genes, as a group, are enriched for pseudogenes. Additionally, Class 1 pairs with the scB copy expressed in at least one SRA library have higher mean K_a/K_s ratios than those of Classes 2–4 (Class 1: 0.65; Classes 2–4: $\leq$0.44), further suggesting enrichment for pseudogenes. Nonetheless, because mean Ka/Ks is below 1 for these Class 1 pairs, and lower than for the 7% of Class 1 pairs for which there is no evidence of scB expression in any SRA libraries (mean = 1.41, median = 0.92; $t = -7.3$, df = 148, $p < 0.001$), it is likely that at least some Class 1 scB genes expressed in SRA libraries are functional. Class 1 had the highest standard deviation in K_a/K_s (**Supplementary Table 6**), consistent with this class comprising a mixture of pseudogenes and functional genes under different selective pressures.

In pairs comprising Classes 2–4, both paralogs are expressed in at least one cluster, with Classes 3 and 4 distinguished from Class 2 by biased paralog expression defined at the stringent threshold of 9:1, and Class 3 and 4 pairs differentiated by whether the bias was unidirectional, with only scA ever being dominant (Class 3), or whether the bias was reciprocal, with scA and scB dominant in different RCCs (Class 4; **Table 1**). Whereas the majority of Class 0 and Class 1 pairs are from proximal, tandem or young transposed duplicates, the majority of pairs in Classes 2–4 were produced by WGD or older transpositions (**Figure 5**). This pattern was most pronounced for Class 3 (**Figure 5**). Class 4 pairs comprised by far the smallest number of pairs (only 1.5–5.6% of pairs among duplication types), but showed the same pattern, with WGD and older transposed pairs having a larger percentage representation than the other SSD types.

Expression classes aggregate data across all RCCs to summarize expression patterns, but do not provide information about paralog pair behavior in individual RCCs. Different duplication types also showed very different percentages of pairs exhibiting bias in individual RCCs, with proximal, tandem, and younger transposed classes all showing greater levels of bias than WGD and older transposed pairs (**Figure 6**). Homoeologs from alpha WGD pairs are the least likely to show expression bias. The fraction of biased pairs also varied by RCC, with root cap clusters generally exhibiting the least bias across all types of duplicates (**Figure 6**).

Pairs Showing Extreme Reciprocal Paralog Expression Bias

Class 4 pairs are defined by extreme reciprocal expression biases ("reciprocal fixation") of paralogs across clusters, and are the

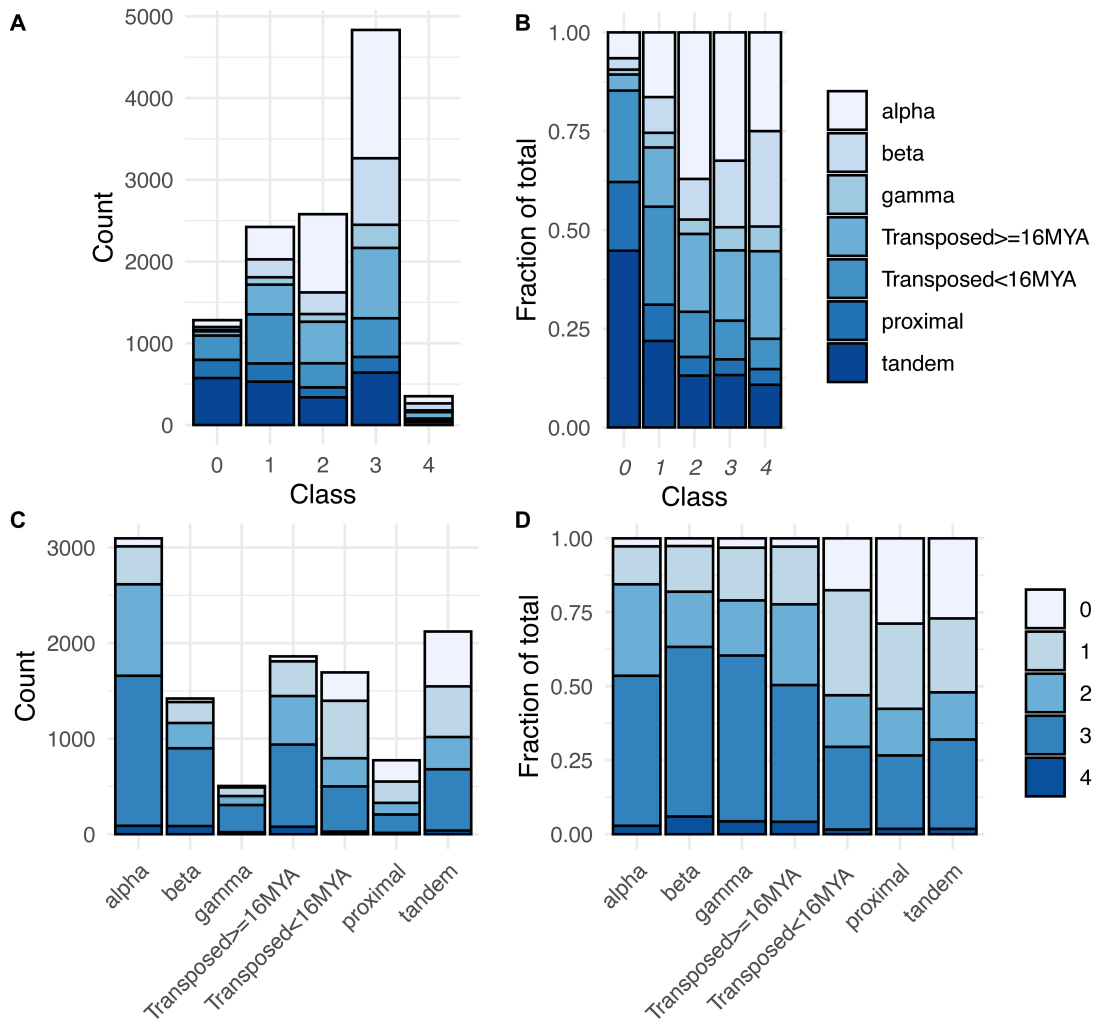


FIGURE 5 | Distribution of expression classes and duplication types for gene pairs. **(A)** Counts of gene pairs by expression class, broken down by duplication mechanism. **(B)** Fraction of gene pairs in each expression class produced by each duplication mechanism. **(C)** Counts of gene pairs by duplication mechanism, broken down by expression class. **(D)** Fraction of gene pairs from each duplication mechanism assigned to each expression class.

most likely cases of expression subfunctionalization following duplication. To determine the degree to which cell-level data provided greater resolution to detect reciprocal fixation, we tallied the number of Class 4 genes: (1) between bulk tissues (root vs. leaf SRA libraries); (2) across the 9 superclusters, roughly representing cell types; (3) across the 36 RCCs, further subdividing putative cell types into cell states; and (4) between individual cells of the 36 RCCs (**Figure 1**).

The number of cases of reciprocal paralog expression bias identified using the single cell data was three times greater than those identified in the bulk tissue comparison (403 vs. 124; **Figure 7**). Most of these cases (352) were identified at the level of the 36 RCCs, suggesting that in root tissue, expression subfunctionalization occurs more frequently among cell states within a cell type than among cell types, with the caveat that the 9 superclusters may include more than one cell type, and that such heterogeneity could reduce the estimated number of Class 4 pairs.

Within the 9 superclusters, only 56 gene pairs exhibited reciprocal fixation at the level of single cells (an over-representation of cells exhibiting significant bias favoring one copy in some cells, and the other copy in other cells, with few or no cells co-expressing both equally). This is likely an undercount of the true number, however, due to the low read count per cell. Of the 87.5 million possible cell × gene pair comparisons (7,519 cells × 11,631 gene pairs), 19.2 million had non-zero read counts, but of these, 17.5 million had counts that were nonetheless too low to detect bias at our threshold of 9:1.

In contrast to 56 significantly non-overlapping gene pairs, 1,363 gene pairs exhibited significant overlap at the level of single cells (cells expressing one copy were significantly more likely to express the other copy as well). Thus, at the single cell level, paralogs appear to be coexpressed 24.3-fold (1,363/56) more often than not. Alpha WGD duplicates were the most likely to exhibit significant overlap at the level of single cells,

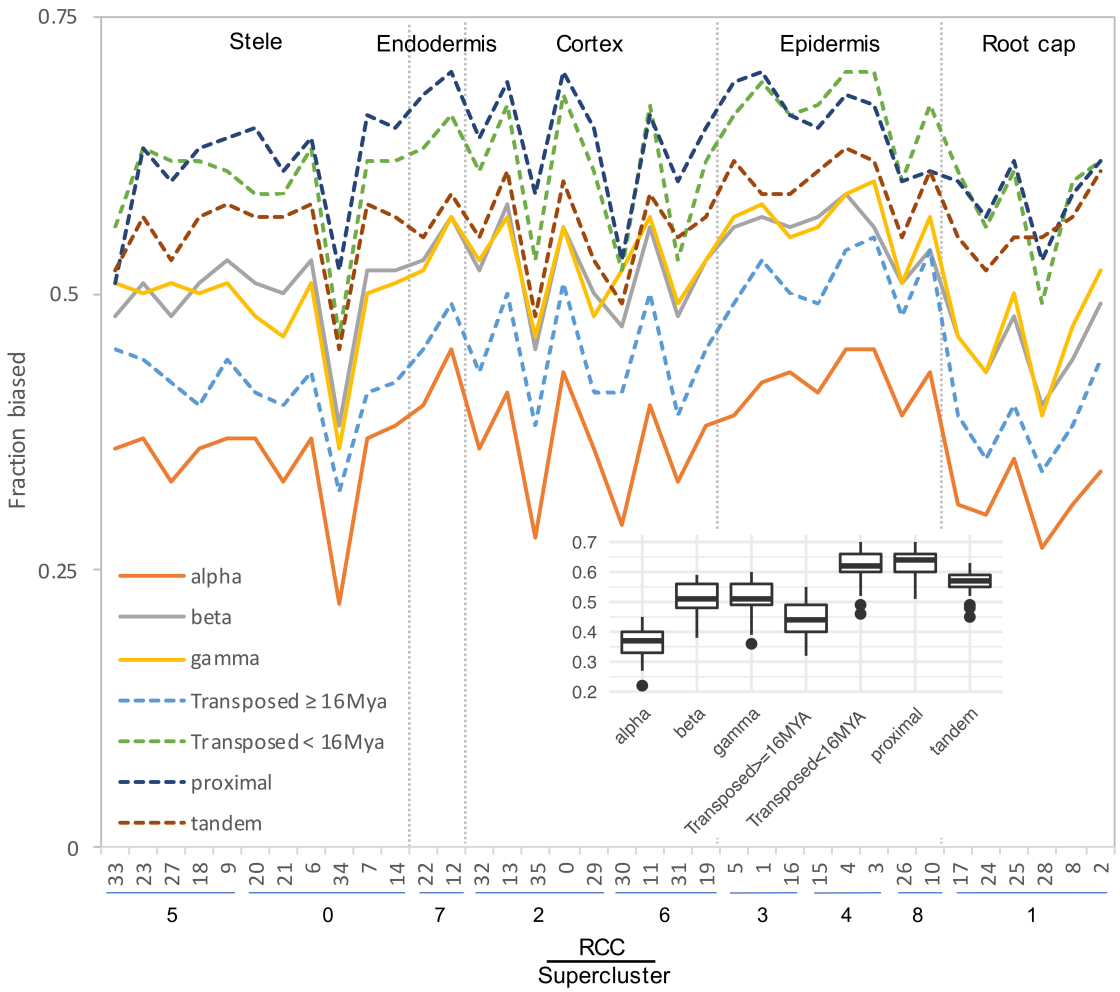


FIGURE 6 | Fraction of gene pairs exhibiting paralog expression bias by RCC. RCCs are arranged by supercluster and ordered roughly from the interior (stele) to the exterior (epidermis) and tip (root cap) of the root. Each line indicates the fraction of gene pairs from the specified duplication mechanism that exhibit expression bias per RCC. For ease of comparison, whole genome duplications are shown with solid lines, and small scale duplications are shown with dashed lines. Individual gene pair by RCC combinations that lacked sufficient read depth to detect bias were excluded from the analysis. Inset, box plot summarizing the distribution of bias fractions by duplication type.

whereas gamma WGD duplicates were the most likely to exhibit significant non-overlap (**Table 2**).

Notably, Class 4 pair genes at all four levels of organization (bulk tissue, cell types, cell states, and single cells) are enriched for extracellular functions (e.g., extracellular region, apoplast, cell-cell junction), and each is also enriched for some aspect of the cell periphery (e.g., cell wall, plasma membrane) ($p < 0.05$; Holm-Bonferroni; **Supplementary Table 3**). Thus, reciprocal fixation of expression, suggestive of partitioning of function, appears to occur preferentially among paralogs functioning at the cell surface. Beyond this commonality, however, GO enrichment analysis suggests that paralogs exhibiting extreme reciprocal fixation at the bulk tissue level differ functionally from those exhibiting reciprocal fixation at finer levels of resolution. At the bulk tissue level, Class 4 pair genes are preferentially involved in lipid metabolism and vesicle trafficking (exocyst), whereas at the supercluster, RCC, and single cell levels, Class 4 pair genes are preferentially involved in cell wall modification (e.g., cell wall organization or biogenesis, hemicellulose metabolic process) and response to stress (e.g., response to oxidative stress, response to toxic substance; **Supplementary Table 3**).

Concerted Divergence of Paralogous Genes

Overall, WGD gene pairs were more likely to exhibit reciprocal fixation across the 36 RCCs than were SSDs ($\chi^2 = 22.1$, $p < 0.001$). In total, 195 out of 5,018 WGD pairs (3.9%) were assigned to Class 4, compared to 157 out of 6,449 SSD pairs (2.4%).

It has been proposed that WGD facilitates functional differentiation by simultaneously duplicating entire gene networks, thereby providing the raw material for "concerted divergence" (Blanc and Wolfe, 2004) of whole biological

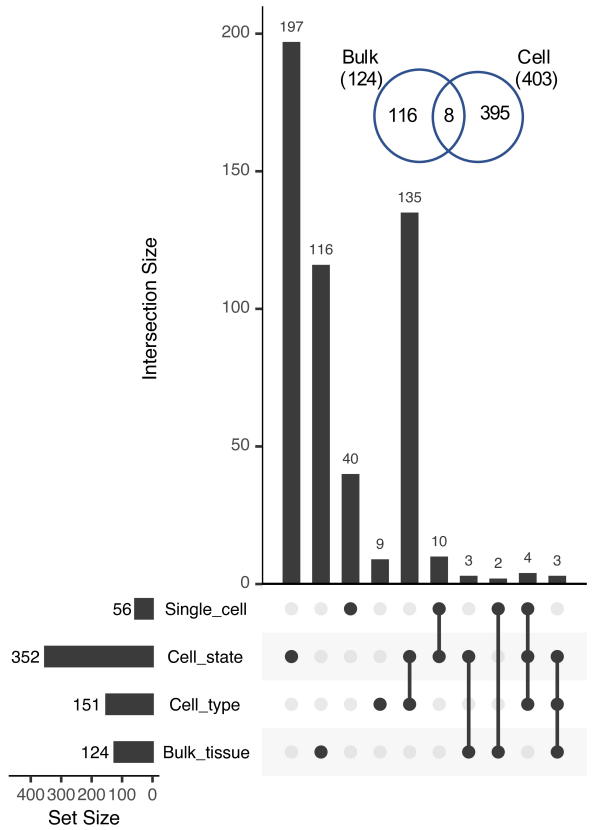


FIGURE 7 | UpSet plot of intersections among Class 4 genes identified at different levels of organization. Inset, Venn diagram showing the overlap between Class 4 pairs identified in bulk tissue (Bulk) vs. those identified at sub-tissue levels (Cell; including cell type, cell state, and single cells).

pathways. We looked for evidence of this phenomenon, focusing on the genes duplicated by the alpha WGD. Out of a total of 3,096 alpha gene pairs, 176 genes in 88 pairs belonged to Class 4, the class most consistent with expression subfunctionalization within roots.

For each of these 88 Class 4 alpha WGD pairs we calculated an expression ratio (scA/total) for each of the 36 RCCs. Then, for each alpha pair, we calculated correlation coefficients with every other alpha pair based on these 36 expression ratios. To the extent that two alpha pairs have diverged in concert, we would expect their expression ratios to be either positively correlated ($r > > 0$) if scA from both pairs are coevolving, or negatively correlated ($r < < 0$) if scA from one pair is coevolving with scB from the other pair. For alpha pairs that are diverging independently of each other, we expect to see no correlation ($r = 0$). By this approach, 33 of the 88 pairs formed a distinct set with significantly correlated expression ratios ($p < 0.05$; **Figures 8A,B**), suggesting concerted divergence into two separate networks. Several additional, smaller clusters were evident as well.

Within the 33 pair cluster, paralogs have diverged in expression such that one copy of each pair is preferentially expressed in the stele and the other copy is preferentially expressed in the epidermis (**Figure 8C**). Paralogs exhibited a range of biases in intervening layers. The stele-dominant copy of each pair is also dominant in columella cells and one population of meristematic cells in the root cap, whereas the epidermis-dominant copy is weakly dominant in a second population of dividing meristematic cells. As expected, pairwise correlation coefficients of expression profiles were higher within the two diverged gene sets (epidermis-biased and stele-biased) than among the Class 4 alpha pairs not in either gene set, or between the epidermal and stele gene sets (**Figure 8D**), consistent with the possibility that the two gene sets have diverged in concert and are acting as separate functional modules.

TABLE 2 | Counts of overlap and non-overlap of paralog expression in individual cells.

Duplication type	Overlap	% of total	Non-overlap	% of total
alpha	593	18.64%	13	0.41%
beta	170	11.72%	5	0.34%
gamma	64	12.28%	4	0.77%
Transposed ≥ 16 MYA	245	13.16%	13	0.70%
Transposed < 16 MYA	103	6.06%	7	0.41%
proximal	39	4.97%	2	0.26%
tandem	149	7.00%	12	0.56%

The 66 genes in the cluster are not enriched for any GO terms or protein domains, though several genes are involved in calcium signaling. Similarly, neither set of 33 genes in the two separate co-expressed networks is enriched for GO terms, and both homoeologs have equivalent annotations in most cases. Thus, there is no obvious functional differentiation discernable at the level of gene ontology between homoeologs in the two networks.

However, 13 genes in the epidermal network have protein-protein interactions annotated in the InTact database[2], and two of these genes encode proteins that interact directly: a MYB transcription factor encoded by AT2G31180 and a calcium-sensing calmodulin protein encoded by AT4G14640. A third gene (AT4G30560) encodes a calmodulin-regulated ion channel that interacts indirectly with the calmodulin protein via a protein kinase intermediary (AT4G04570), suggesting a gene module involved in calcium-dependent ion trafficking functioning in the epidermis. No equivalent interactions are annotated among the stele-biased genes, and no genes from the epidermally-biased co-expression network directly interact with genes from the stele-biased network. Collectively, these observations suggest that the two homoeologous networks have diverged in concert, both spatially and functionally.

Correspondence of RCC Expression Classes With Functional Divergence of Paralogs

Expression and function are generally synonymized in discussions of gene evolution (e.g., Panchy et al., 2019). For gene pairs with biased paralog expression profiles, we therefore

[2]https://www.ebi.ac.uk/intact/

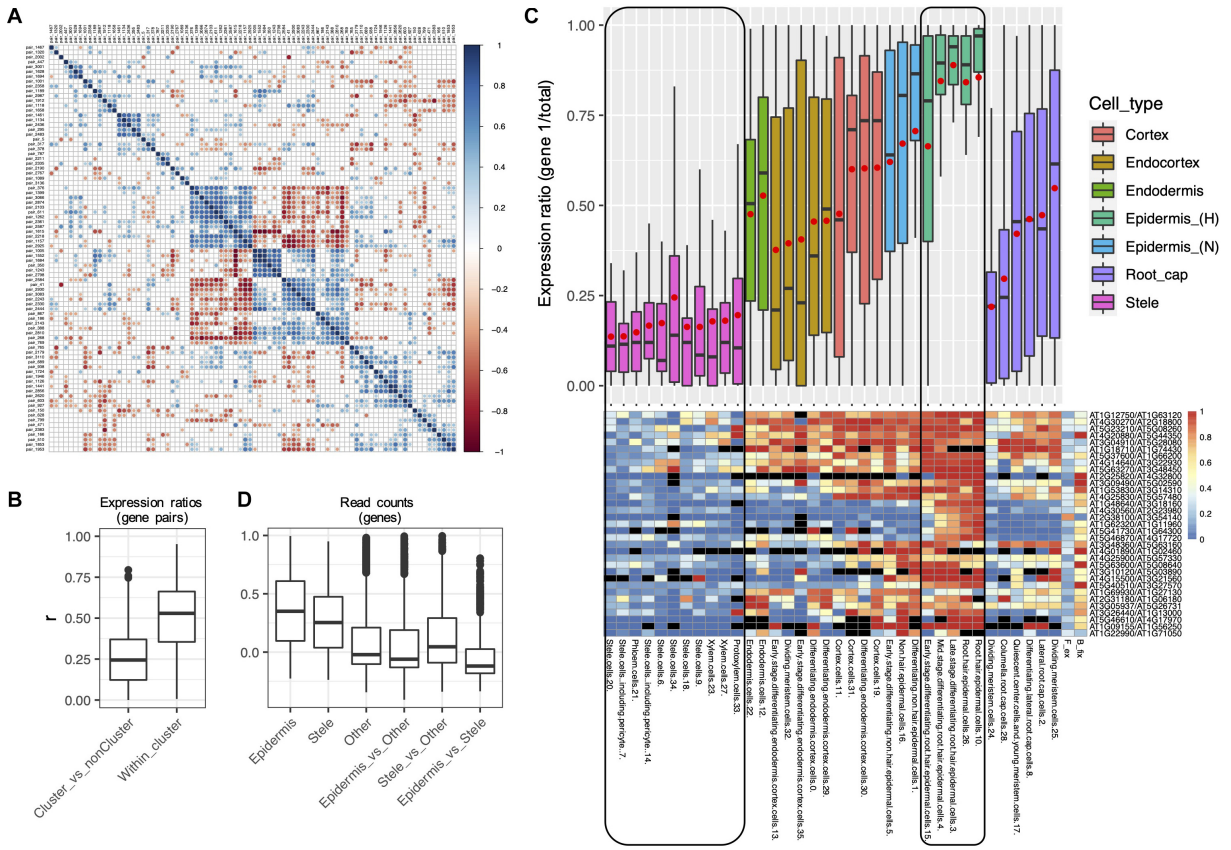


FIGURE 8 | Evidence for concerted divergence among alpha homoeologs. **(A)** Correlogram of pairwise correlation coefficients of expression ratios for all 84 alpha Class 4 gene pairs. Each row or column represents a single gene pair. Significant positive pairwise correlations (scA from both pairs show similar patterns of dominance/non-dominance) are shown in blue, and significant negative correlations (scA from one pair shows a similar pattern as scB from the other pair) are shown in red. Non-significant correlation coefficients are shown in gray. A cluster of 33 significantly correlated pairs is visible in the middle of the plot, in which each pair has diverged in expression in a correlated manner with each other pair, suggesting concerted divergence. **(B)** Pairwise correlation coefficients (r) of homoeolog expression ratios within the cluster of putative concerted divergence (Within cluster), and between this cluster and all other Class 4 alpha pairs (Cluster_vs_nonCluster). **(C)** Boxplot and heatmap showing the distribution of expression ratios (scA/total) for the 33 gene pairs comprising the largest cluster of alpha gene pairs exhibiting putative concerted divergence. Expression ratios are shown for each of the 36 RCCs (cell states), and these are clustered and color coded by cell types. A common color was assigned to the two "Stele" superclusters and to the two "Epidermis (H)" superclusters to emphasize the partitioning of expression between these cell types. Red points in the boxplot indicate mean values. Heatmap colors indicate raw expression ratios (gene 1/total) without application of Wilson tests. **(D)** Pairwise correlation coefficients of read counts by RCC among the homoeologs within the putative cluster of concerted divergence showing epidermal bias (Epidermis) or stele bias (Stele), for all other Class 4 alpha genes (Other), and between each of these gene sets.

assessed the correspondence of different expression patterns with other measures of functional differentiation.

Hanada et al. (2009) categorized 492 *Arabidopsis* gene pairs as showing low, medium, or high morphological diversification on the basis of knockout phenotypes for one or both paralogs. Of these, only 94 were included among the Wang et al. (2013) duplicate pairs, of which one or both copies were expressed in roots for 90 (**Supplementary Table 3**). Despite the small number of pairs in both data sets, the distribution of pairs exhibiting or not exhibiting some degree of morphological variation differed significantly by expression class ($\chi^2 = 12.1$, $p = 0.02$). Unlike all other classes, the majority of Class 2 pairs (no biased paralog expression) exhibited no morphological divergence, and no Class 2 pairs exhibited high morphological divergence (**Figure 9A**). If this small sample is representative, the results suggest that pairs of genes with unbiased paralog expression can tolerate the loss of expression from one paralog, suggesting that the two paralogs are not currently maintained either by dosage constraints or by essential functional differences between the two paralogous proteins.

In contrast to Class 2, only around 10% of Class 1 pairs were in the no morphological diversification category, suggesting that paralogs of Class 1 pairs in which the non-root paralog is not a pseudogene have diverged in function. Classes 3 and 4 were intermediate between Classes 1 and 2 (**Figure 9A**). Like Class 1, Classes 3 and 4 both partition paralog expression, but in these classes partitioning is within the root, between different RCCs. Shared root expression could suggest less functional differentiation between paralogs; moreover, many pairs have both paralogs expressed in RCCs other than those fixed for one

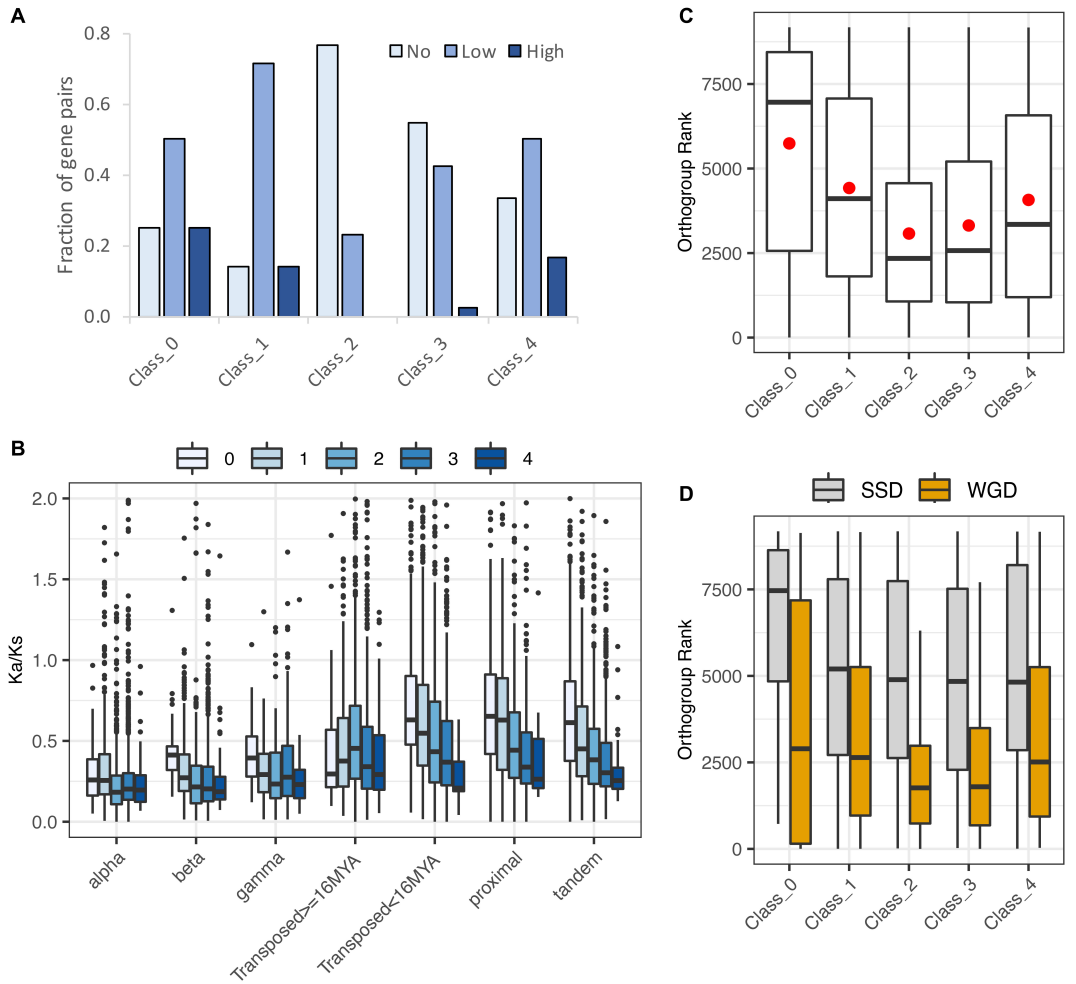


FIGURE 9 | Evidence for functional divergence and gene dosage sensitivity. **(A)** Levels of morphological divergence (Hanada et al., 2009) by expression class. **(B)** K_a/K_s by duplication mechanism and expression class. K_a/K_s values were cut off at 2 for clarity of presentation, but extended beyond this value in all cases. **(C)** Correspondence of expression classes with dosage sensitive gene families ("orthogroups": Tasdighian et al., 2017). Lower orthogroup rank indicates greater overall dosage sensitivity. Orthogroup rank by expression class for 5,387 duplicate pairs. **(D)** Orthogroup rank by expression class separately for WGD and SSD duplication types.

paralog, making the pair's behavior in those RCCs more like Class 2. The higher proportion of Class 4 pairs showing some degree of morphological diversification suggests that reciprocal fixation – the near-exclusive use of a different paralog in different RCCs – may be more indicative of a shift in function between paralogs than consistent dominance of one paralog (Class 3).

The ratio of replacement to silent nucleotide substitutions (K_a/K_s) is understood to be a measure of functional divergence. For all duplication types and expression classes, K_a/K_s distributions were strongly skewed, with median and mode less than 1.0 (**Figure 9B** and **Supplementary Tables 4**, 6), indicating that most gene pairs are evolving under purifying selection, as expected. As reported by Qiao et al. (2019) for 141 phylogenetically diverse plant genomes, average K_a/K_s varied with duplication type, with WGD classes showing lower K_a/K_s than tandem, proximal, and young transposed pairs. Variances among both duplication types and expression classes differed significantly ($p < 0.001$), suggesting differences in the number of pairs showing positive selection or containing an expressed pseudogene. Class 1 had the highest standard deviation in total and for all duplication types except proximal SSDs, suggesting that this class includes either pseudogenes, genes under positive selection, or both. Class 4 had the smallest standard deviation (**Supplementary Table 6**). The large variance for Class 1 is also consistent with its having the highest percentage of morphologically divergent paralog pairs as classified by Hanada et al. (2009).

The fact that Class 4 has the lowest average K_a/K_s, and the smallest variance around the mean seems to indicate uniformly strong purifying selection. The relatively low level of amino acid replacements between Class 4 paralogs, coupled with the functional divergence (see above), suggests that these pairs may be diverging in their promoter regions (sequence and/or accessibility).

Dosage Sensitive Gene Families

Many duplicate gene pairs are thought to be preserved not by functional divergence, but to maintain stoichiometric relationships in protein complexes and interacting networks following whole genome duplication; the same constraints lead to the loss of paralogs following SSD (e.g., Freeling, 2009; Birchler and Veitia, 2010, 2012, 2014). Tasdighian et al. (2017) ranked 9,178 core gene families ("orthogroups") across 37 angiosperm based on how strongly paralogs have been retained in duplicate following WGD vs. SSD; gene families with high levels of WGD retention and minimal SSD retention are defined as "reciprocally retained," and strong reciprocal retention is considered a hallmark of selection to maintain dosage balance. 6,308 Wang et al. (2013) pairs occur among these 9,178 orthogroups; 853 of these pairs had paralogs assigned to two different orthogroups, and an additional 68 were not assigned to expression class because one gene model was obsolete; these were excluded, leaving 5,387 pairs with one or both members assigned to an orthogroup and class (**Supplementary Table 3**). We divided the 5,387 pairs by expression class and determined the "orthogroup rank" as in Tasdighian et al. (2017), wherein orthogroups with the highest reciprocal retention are ranked the highest (i.e., smallest numerical rank). Class 2 pairs were the most dosage-sensitive, significantly more highly ranked than other classes (Class 2 vs. Class 3: $z = -2.22$, $p = 0.029$; Dunn's test of multiple comparisons with Benjamini-Hochberg correction; all other comparisons significant at $p \leq 0.0001$; **Figure 9C**). This is consistent with the expectation that if dosage balance is responsible for preserving both paralogs, neither paralog is likely to be strongly biased in its expression. Other classes showed less evidence of dosage-based constraints, with Class 1 pairs showing the least effect (highest orthogroup ranks), consistent with its paralogs being expressed in different organs. Although Class 3 and 4 pairs have some RCCs fixed for one paralog, other clusters have both paralogs expressed, so it is possible that dosage balance is necessary in some cell types but not in others, where divergence of expression pattern ("function") is occurring.

Dividing the 5,387 pairs into WGD and SSD duplication types (**Figure 9D**) showed the expected result that SSD pairs are less affected by dosage constraints than are WGD pairs. Alpha WGD duplicates showed the same overall pattern of classes as did the full dataset and the WGD duplicates combined (data not shown).

Expression Patterns of Duplicate Pairs by Subgenome

The loss of duplicated genes following polyploidy – fractionation (Langham et al., 2004) – is often unequal across homoeologous subgenomes, and such biased fractionation is thought to begin with unequal expression of homoeologs (e.g., Steige and Slotte, 2016). Schnable et al. (2012) assigned *Arabidopsis* homoeologs from 817 *Arabidopsis* alpha WGD pairs to two putative homoeologous subgenomes, A and B, with A being the dominant subgenome, characterized by lower rates of gene loss and higher expression of remaining genes. We compared subgenome assignments with our scA/scB classification based on expression of Wang et al. (2013) alpha WGD homoeologs in RCCs, and found representation of pairs from all our classes, with 9 Class 0 pairs, 82 Class 1 pairs, 240 Class 2 pairs, 377 Class 3 pairs, and 23 Class 4 pairs (86 pairs from Schnable et al. (2012) were not included in the Wang et al. (2013) assignments; **Supplementary Table 3**). We wished to determine whether the paralog with higher expression in RCCs most commonly was from the dominant subgenome. For this, Class 0 (no root expression) and Class 2 (no extreme expression bias) pairs were irrelevant, whereas Class 1 may have bias, and Classes 3 and 4 must have at least one biased context. Of the 453 pairs in the Schnable et al. (2012) assignments for which we did see bias in one or more contexts (**Supplementary Table 3**), the scA/scB assignments based on dominance observed in our expression data were consistent with the A/B subgenome dominance assignments 261 times, which is significantly more than expected by chance ($p = 0.0007$; binomial test). Of the three classes, Class 1 was most concordant, with approximately 4 times the number of pairs agreeing than disagreeing, while Classes 3 and 4 both had roughly only 1.25 times more cases of agreement than disagreement, again a statistically significant difference ($p = 0.003$; Fisher's exact test). Though the majority of Class 1 scB paralogs are expressed in other tissues (**Supplementary Table 5**), they are expressed at lower levels and under fewer conditions than their scA counterparts (**Supplementary Figure 4**). Their over-representation in the non-dominant subgenome, therefore, is consistent with the hypothesis that genome dominance (fractionation bias) is driven by expression level bias. The average F_{ex} value for pairs in agreement with the subgenome assignments was 0.36 compared to 0.31 for those in disagreement.

We also asked if the 33 alpha pairs representing potential concerted divergence (**Figure 8**) have partitioned expression by subgenome. The stele-dominant set includes eight homoeologs from subgenome A and five from subgenome B, and the epidermis-dominant set includes five homoeologs from subgenome A and seven from subgenome B. 41 of the 66 genes in the two sets genes were not assigned to subgenome by Schnable et al. (2012). This lack of subgenome assignment for most genes makes it difficult to assess patterns of subgenome partitioning, but the mixed representation in each pathway suggests that the two sets most likely diverged in concert after the alpha WGD, enlisting genes from both subgenomes, rather than in the progenitors of the polyploid.

Shared and Unique Transcriptomes of Cell Clusters and the Evolutionary Polarity of Duplicate Pair Expression

The 22,669 genes that are expressed in at least one of the 36 RCCs include 7,653 genes, comprising 33.8% of the overall root transcriptome and well over half of the transcriptomes of some RCCs, that are expressed in all 36 root cell clusters (RCC-u genes; **Figure 4** and **Supplementary Table 3**). We found that over 99% of RCC-u genes are expressed in at least one non-root SRA dataset (data not shown). Among genes expected to belong to the RCC-u class are genes that are expressed in all cells of the plant (both root and non-root), as well as genes expressed in all root cells but not in all cell types of other plant tissues. A total

TABLE 3 | Breakdown by expression class of 4,599 Wang et al. (2013) paralog pairs ubiquitously expressed in root cell clusters (RCC-u gene pairs).

Class	Count	RCC-u mixed	RCC-u both
0	0	0	0
1	505	505	0
2	1,174	327	847
3	2,780	2,126	654
4	140	89	51

of 3,505 RCC-u genes (45.7% of RCC-u genes) were among the set of 4,577 genes expressed in all 11 *Arabidopsis* tissues studied by Cheng et al. (2017) Thus, nearly half of RCC-u genes may be ubiquitously expressed in plant cells.

RCC-ubiquitous genes are thus an interesting class to consider because their presence in all RCC transcriptomes suggests that their expression is indispensable in all root cell types and states, and, for over 45%, perhaps in all cells. This suggests that prior to any duplication – whether WGD or SSD – each of these genes was expressed as a single copy in all RCCs, and therefore that immediately after duplication both paralogs were RCC-u. Under these assumptions, pairs in which only one paralog is ubiquitously expressed in root cells are inferred to represent evolutionary losses of expression of the non-RCC-u paralog, rather than gain of expression in some RCCs leading to ubiquitous expression of the RCC-u paralog. We further assumed that, immediately following duplication, the two paralogs were expressed at similar levels.

Among the 7,653 RCC-u genes are 4,538 genes belonging to 4,639 Wang et al. (2013) gene pairs (a gene can belong to more than one pair due to nested duplications). 40 of these gene pairs include one gene that is RCC-u and one that is an obsolete gene model; consequently, these pairs are not assigned to an expression class, leaving 4,599 Wang et al. (2013) RCC-u gene pairs assigned to an expression class. Only 1,552 pairs retained both genes as RCC-u, and among these, only 847 pairs did so without significant expression bias (Class 2; **Table 3**). Thus, both copies retain the putative ancestral expression profile for only 18.4% (847/4,599) of RCC-u pairs. 705 pairs retained both genes as RCC-u but showed biased expression in one or more RCCs (Classes 3 and 4). The remaining 3,047 pairs include one copy that is RCC-u and one that is not, indicating changes in expression in at least one copy from the ancestral profile. For 327 pairs, these shifts were subtle, as we did not detect significant bias (Class 2). In the remaining cases (2,720 pairs), the shift from the hypothesized ancestral condition of both paralogs being RCC-u was more dramatic, including the 505 pairs for which the second paralog was not expressed in any root cell type (Class 1; **Table 3**).

These counts are further broken down by duplication mechanism in **Supplementary Table 8**. For alpha duplicates, there were 1,685 pairs with one or both genes RCC-u. Of these, in 294 pairs both homoeologs were RCC-u but show significant bias in at least one cluster (17.4%; Classes 3 and 4), and 949 pairs (56.3%) have only one RCC-u copy with varying degrees of bias in clusters where both homoeologs are expressed (**Supplementary Table 8**). Conversely, 442 pairs (26.2%) show no evidence for shifts from the ancestral state (both copies are RCC-u with no bias [Class 2]). This is the highest fraction of pairs with both copies retaining the ancestral expression pattern of any duplication type. Gamma and beta WGD duplicates exhibited the next highest fractions (16 and 15.1%, respectively), and the SSD duplicates had the lowest. This suggests that WGD duplicates are more constrained to retain their ancestral expression patterns, perhaps due to dosage constraints.

Representation in the RCC-u class itself also varied with duplication type, with over-representation of WGD duplicates expressed across all RCCs relative to the total fraction of WGD duplicates in the genome (**Supplementary Figure 5**). In contrast, SSD duplicates, other than older transposed duplicates, were under-represented relative to their representation in the genome as a whole, more like genes lacking a duplicate in the Wang et al. (2013) set and for which one paralog has presumably been lost after duplication ("singletons" in **Supplementary Figure 4**).

Evolutionary shifts in expression of paralog pairs were also observed among the remaining 59.6% of pairs comprising the cumulative root transcriptome that are expressed in 1–35 RCCs (non-RCC-u pairs). Their lack of ubiquitous expression in root cell types/states makes it more difficult to determine whether expression of a given gene in a particular RCC represents the ancestral state or the derived state, and thus to hypothesize the ancestral condition for a paralog pair with only one member expressed in an RCC. If we assume that at the time of a duplication, both paralogs retained the expression pattern of the single copy gene progenitor (see below for discussion of this assumption), then pairs with unbiased expression of both paralogs or neither paralog in a given RCC are most readily explained as retaining the expression state of their single copy progenitor, since two independent gains or losses of expression would need to be hypothesized otherwise.

Pairs with only one paralog expressed in any RCC (Class 1), or with biased expression of one paralog (Classes 3 and 4), could equally parsimoniously be inferred to have gained or lost expression of one paralog in RCCs for which the pair shows fixation. In either case, however, an evolutionary shift in expression from the inferred ancestral condition is involved. The total number of RCCs showing fixation of a paralog pair in the non-RCC-u class (i.e., the number of red or blue cells in a heatmap of non-RCC-u pairs, similar to the examples shown in **Figure 2**) was 22,897; this was 45% of the total number of number of heatmap cells (non-RCC-u paralog pairs × 36 RCCs; **Supplementary Table 9**).

Expression Differences Among and Within Cell Types and Cell States

In addition to assessing the behavior of gene pairs following their duplication by WGD or SSD, we were interested in exploring differential responses of different cell types/states to gene duplication. The 9 superclusters identified by Ryu et al. (2019) in some cases aggregated known cell types (e.g., Ryu supercluster 1, in which cells of the quiescent center are grouped with root cap cells; also Ryu superclusters 0, 2, and 5). These presumably artificial superclusters were disaggregated in the 36

RCC data studied here (e.g., **Figures 2**, 3, 6). This resulted in the identification of some extreme reciprocally biased expression patterns between RCCs of the Ryu et al. (2019) superclusters (**Figure 6**). For example, there were 32 reciprocal fixations within Ryu supercluster 5, with protoxylem cells being particularly differentiated in their expression pattern relative to other cell types grouped in the same supercluster.

Ryu et al. (2019) studied the developmental trajectories of root hair cells, as well as other cell types, and those cell developmental states are also included in the 36 RCC data, allowing us to search for differential responses to duplication among developmental states of more confidently identified cell types: non-hair epidermal cells (three RCCs of supercluster 3), root hair epidermal hairs (three RCCs of supercluster 4), late stage and mature cortex cells (four RCCs of supercluster 6), mature endodermis cells (two RCCs of supercluster 7), and a second root hair supercluster (two RCCs of supercluster 8). No cases of reciprocal fixation were found among states belonging to superclusters 4, 7, or 8, but a beta WGD pair (AT3G13750/AT5G56870) was found to be reciprocally fixed for states within both superclusters 3 and 6; the alpha pair AT3G18950/AT1G49450 (row 12 in **Figure 2A**) and the young transposed pair AT5G39580/AT5G64100 were found to be reciprocally fixed for states within supercluster 6.

In addition to the small number of reciprocal shifts in extreme paralog expression bias among states of the same cell type, there were many other instances where the expression patterns of paralog pairs were not homogeneous across states of a given homogeneous cell type (e.g., rows 1, 2, 7, 8, 11, and 12 in **Figure 2A**). For example, there were 87 alpha WGD pairs that had no expression of either paralog in one RCC of non-hair epidermal cell supercluster 3, fixation of one paralog in a second RCC of the supercluster, and balanced expression of both paralogs in the third RCC (**Supplementary Table 7**). Overall, for supercluster 3, 35% of alpha WGD pairs (1,119/3,181) had non-homogeneous expression across its three RCCs; for all duplication types, the average was 30% for this supercluster. In all five of these putatively homogeneous cell type superclusters, the three WGD classes and the older transposed duplication class had higher percentages of non-homogeneous expression; the actual percentages were roughly correlated with the number of RCCs in a supercluster – finer division resulted in greater heterogeneity – but superclusters with the same number of RCCs differed from one another (e.g., superclusters 3 and 4, each with three RCCs, had overall heterogeneity percentages of 30 and 23.3%, respectively), suggesting cell type-specific patterns of paralog expression during differentiation.

To explore differential responses of RCCs to different types of gene duplication, we calculated a "paralog expression retention" (PER) index, defined as the number of pairs expressing both paralogs in a given RCC divided by the number of pairs with at least one paralog expressed in the RCC, and compared this value across RCCs for all pairs except those for which both paralogs were classified as RCC-u. There was significant variation in PER by both RCC (χ^2 = 4285.8, df = 35, $p < 0.001$) and duplication type (χ^2 = 615.1, df = 6, $p < 0.001$), with WGD and older transposed duplicate pairs showing higher rates of paralog retention (**Figures 10A,B**). Variation in PER across RCCs was likely due, in part, to differences in sequencing depth per RCC (RCCs with fewer total reads are more susceptible to dropout). On the assumption that sequencing depth has the same relative effect on PER for different duplication mechanisms in all RCCs, we adjusted for differences in sequencing depth by calculating the differential retention ratio, PER_{wgd}/PER_{ssd} for each RCC. Because of our specific interest in the alpha WGD we also calculated PER_{alpha}/PER_{tandem} (**Figure 10C**). These two ratios were similar (**Figure 10D**). Variation in differential retention ratio across RCCs is likely due to differences in cell biology that reflect the degree to which the genes whose expression comprises the transcriptome are dosage sensitive; higher PER of WGD pairs is expected given their greater representation among dosage sensitive gene families (Tasdighian et al., 2017; Qiao et al., 2019).

DISCUSSION

Against the backdrop of constant and ongoing single gene duplications (Lynch and Conery, 2000), the genome of *A. thaliana* bears the legacy of multiple polyploidy events, dominated by the alpha WGD event, which phylogenomic studies indicate took place 32–43 MYA (Edger et al., 2018). The availability of single cell transcriptome data from *A. thaliana* roots (Ryu et al., 2019) allowed us to assay the effects of this event on gene expression in unprecedented detail, from the dual perspectives of the genes duplicated by the polyploidy event and of cell types that, after millions of years of existence at a lower ploidy, were suddenly presented with the challenges and opportunities of a doubled genome.

Expression Evolution of Alpha WGD Homoeolog Pairs in Root Cell Types and Cell States

For simplicity we have assumed, as others have done (e.g., Blanc and Wolfe, 2004; Panchy et al., 2019) that most alpha WGD pairs initially conserved their ancestral functions and partitioned their expression equally between what are now their two homoeologs. This is a questionable assumption if, as is widely accepted based on biased fractionation of the homoeologous genomes of *Arabidopsis* (Schnable et al., 2012; Woodhouse et al., 2014; Cheng et al., 2018), the duplication was an allopolyploidy event, merging two already differentiated genomes. Although root cell types presumably were conserved between these progenitor species, their transcriptomes no doubt differed from one another, given the effects of mutation and drift even under stable selection (Lynch, 2020). For example, fiber cell transcriptomes differ between the diploid species whose genomes merged to form tetraploid cotton (Yoo et al., 2013; Gallagher et al., 2020). Thus, differences in the expression patterns of alpha WGD homoeologs may be due to "parental legacy" (Buggs et al., 2014; Steige and Slotte, 2016) and not solely to genome duplication. We see some evidence of this in the correlation of our biased expression classes with the homoeologous subgenomes. This is particularly true of Class 1 pairs, many of which may include pseudogenes as the more weakly expressed homoeolog (scB). In the 82 alpha

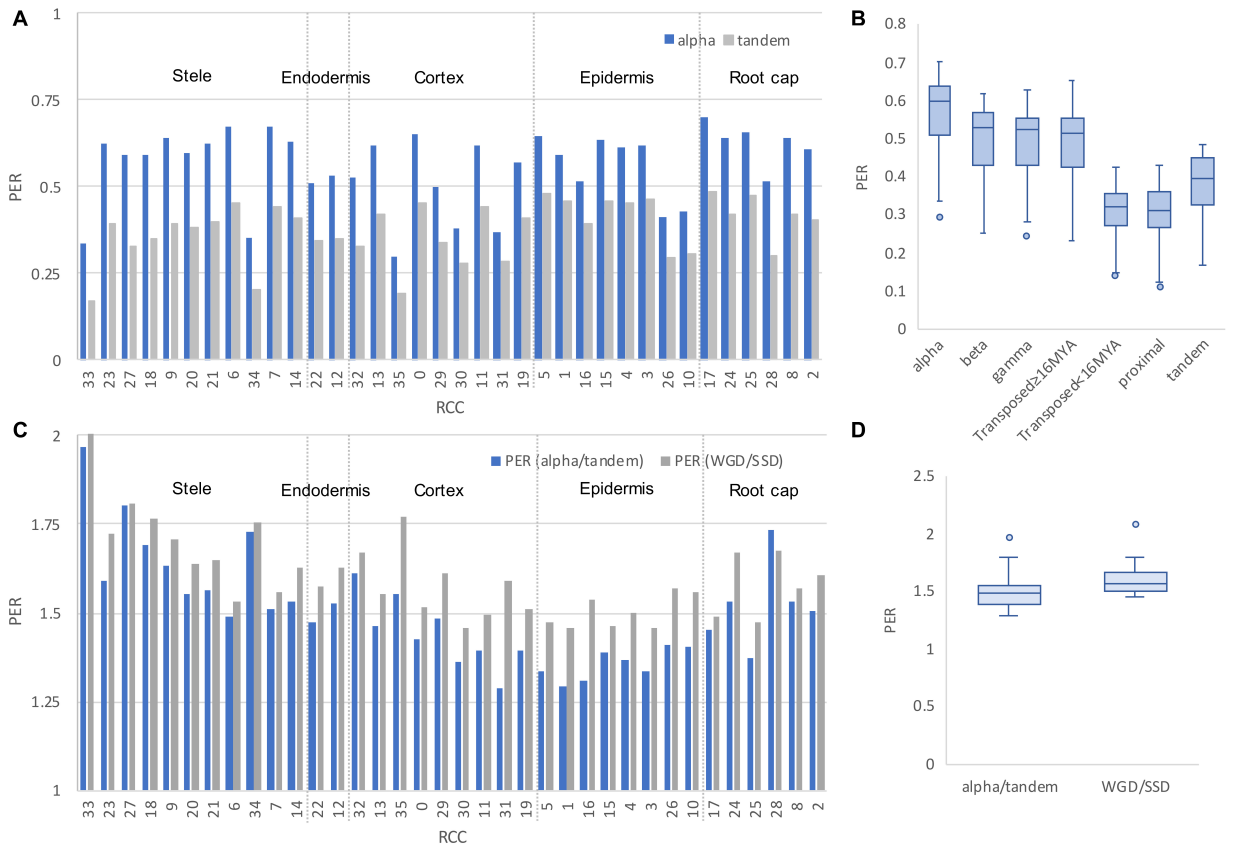


FIGURE 10 | Paralog retention ratio (PER) by RCC and duplication mechanism. **(A)** PER for alpha and tandem duplicates by RCC. The grouping of RCCs by cell type is indicated with dashed lines. **(B)** PER distributions across RCCs by duplication type. **(C)** Differential retention ratios for WGD/SSD and alpha/tandem duplication types. **(D)** Distribution of differential retention ratios (alpha/tandem and WGD/SSD).

WGD Class 1 pairs mapped to subgenomes, we found that scB was four times more likely to be on the more highly fractionated subgenome B.

Nevertheless, it seems likely that, on average, more of the differentiation of alpha WGD homoeologs has occurred subsequent to genome merger, for three reasons. First, the time since polyploidy is at least 3× longer than the time since speciation prior to allopolyploidy (Edger et al., 2018), so even if homoeolog divergence has been proportional to time, more differentiation would be expected to have taken place in the last 30 MY than in the first 10 MY since progenitor speciation. But, second, genome evolution after allopolyploid merger and duplication is thought to be far from the clock-like divergence that might be expected after diploid speciation, beginning with an initial "genomic shock" phase of rapid genetic and genomic change (McClintock, 1984), followed by diploidization that returns the allopolyploid to a more conventional evolutionary rate (Wendel, 2015). Finally, orthologs generally evolve conservatively with respect to function and expression pattern, at least relative to paralogs (Gabaldon and Koonin, 2013).

Even with the simplifying assumption of initially shared expression patterns, however, determining whether the expression of only one paralog in an RCC is the ancestral or the derived state required a second assumption, as no close outgroup is available (e.g., the closest relative to an alpha duplicate pair generally is a beta WGD homoeolog, diverged around 100 MY from either alpha homoeolog; Panchy et al., 2019). We thus focused on gene pairs where at least one member of an alpha homoeolog pair was ubiquitously expressed in all root cell clusters (i.e., pairs that contained at least one RCC-u gene, detectable in all 36 RCCs), reasoning that these genes were likely to be necessary for the functioning of all cell types and cell states in roots. If so, the ancestral condition of the pre-alpha WGD single copy gene was ubiquitous expression in root cells, and the ancestral condition of the two alpha WGD paralogs was equal expression in all RCCs.

With these two assumptions, we looked for departures from the ancestral condition, representing shifts in expression following homoeolog divergence from their common ancestor. Out of 1,685 alpha pairs with at least one RCC-u homoeolog, we found 949 with only one RCC-u homoeolog, including 94 pairs with only one homoeolog expressed in roots (Class 1 alpha WGD pairs; **Supplementary Table 8**). Additionally, we observed 294 alpha pairs with both homoeologs RCC-u, but with biased expression of one or both homoeologs in one or more RCCs

(280 Class 3 and 14 Class 4 pairs, respectively). Thus, in 73.8% of RCC-u alpha WGD pairs, at least one member has changed its expression as homoeologs have diverged.

This logic does not apply to non-RCC-u pairs – those with neither member expressed ubiquitously across the 36 RCCs. However, even without being able to hypothesize the polarity of the change, under the assumption that both paralogs initially were expressed in an unbiased fashion in the same cell immediately after their duplication, we can assess the amount of change, and we found that 72% of alpha WGD non-RCC-u pairs had one homoeolog uniquely or preferentially expressed in at least one RCC. High levels of expression differentiation between homoeologs are not unexpected. For example, in the older (ca. 60 MY) polyploid, *Gossypium raimondii*, over 90% of homoeologs show evidence of sub- or neofunctionalization (Renny-Byfield et al., 2014).

Determination of expression patterns at the level of root cell types/states also provided information on mechanisms of gene retention. The 590 alpha WGD RCC-u gene pairs that express both copies without strong bias (Class 2 RCC-u pairs; **Supplementary Table 8**) are candidates for preservation by dosage balance constraints (Papp et al., 2003; Coate et al., 2016; Tasdighian et al., 2017; Song et al., 2020). More broadly, all 957 alpha WGD Class 2 pairs, including those that are not expressed in all 36 RCCs (non-RCC-u) are also candidates for preservation by dosage balance. Consistent with this, Class 2 alpha WGD genes showed less evidence of functional divergence (**Figure 9A**) than did other classes, were more likely to belong to dosage-sensitive gene families as defined by Tasdighian et al. (2017; **Figures 9C,D**), and, as expected, showed less evidence of positive selection than did SSD pairs or alpha pairs in other expression classes (**Figure 9B**).

Alpha WGD pairs belonging to Classes 3 and 4 provide examples where homoeologs could be maintained by sub- or neofunctionalization. The polarity of expression shifts of the 1,570 alpha WGD Class 3 pairs is unknown, but each includes at least one RCC where there is novel bias subsequent to the divergence of the two homoeologs. This could occur by enhanced expression of one homoeolog or its recruitment to that RCC, the latter being consistent with neofunctionalization. Alternatively, bias could be achieved by diminished expression of the now more weakly-expressed homoeolog (consistent with subfunctionalization). For Class 3 pairs we do not know whether scB is the predominantly expressed paralog in any non-root cell type, which would likely be true if scB is also protected from loss by being essential. For Class 4 pairs, however, there is reciprocal fixation of the two paralogs of 88 alpha WGD pairs (nearly 3% of all alpha pairs) among the 36 RCCs, providing self-contained examples consistent with subfunctionalization (**Supplementary Table 8**) just within root cell types, novel information not available from bulk tissue samples (**Figure 7**). That divergence in expression pattern among these 88 pairs may be functional is suggested by our finding evidence for concerted divergence (Blanc and Wolfe, 2004) of networks between stele and epidermal cells (**Figure 8**).

Responses of Cell Types to the Alpha WGD Event

The basic cell types of the root – xylem, phloem, epidermis, cortex, root hairs, root cap, etc. – are conserved across most plants that have roots, and presumably evolved early in the history of vascular plants (Raven and Edwards, 2001; Kenrick and Strullu-Derrien, 2014; Huang and Schiefelbein, 2015). Immediately following each of the three polyploidy events detectable in genomes of most Brassicaceae, each cell type must also have functioned well enough, through all of the developmental states leading to its mature condition, to allow the plant to survive.

Individual *Arabidopsis* root cell types and states comprising the 36 RCCs evolved their current expression patterns over 30–40 MY of speciation, divergence, and diploidization since the alpha WGD event, each expressing a subset of the homoeologs retained in the genome. Many of the nearly three quarters of the 1,685 alpha WGD RCC-u gene pairs showing evidence of evolutionary shifts in expression in at least one cell type or state (**Supplementary Table 8**) showed such shifts in many or all RCCs (e.g., given adequate statistical power, a Class 1 RCC-u pair would show an expression shift in each of the 36 clusters). Overall, we found that for both RCC-u and non-RCC-u alpha WGD pairs, over one third of all cell type/state × homoeolog pair combinations for which statistically robust measurements could be made showed evidence of expression shifts (**Supplementary Table 9**).

Moreover, each of the 36 cell types and states has had its own characteristic response to gene and genome duplication. For example, excluding RCC-u genes, approximately 60% of alpha WGD pairs had both homoeologs expressed per cluster on average across all 36 RCCs, but in individual cell types/states this value varied from around 30–70% (**Figures 10A,B**). Additionally, the propensity to express both WGD homoeologs as opposed to both SSD paralogs differed among cell types/states. Specifically, across all cell types/states and gene pairs, both homoeologs of alpha WGD RCC-u pairs were expressed about 1.5× more frequently than were both paralogs of tandem SSD pairs, but varied from 1.3× to nearly 2× in different RCCs (**Figures 10C,D**). Variable responses of different cell types/states were also seen clearly in levels of biased expression of the two paralogs of duplicate pairs in the 36 RCCs, with biased expression of alpha WGD pairs averaging around 40% across all RCCs, but varying from less than 25% to nearly 45% showing bias depending on the cell type/state (**Figure 6**). We attribute this variation to functional differences between cell types/states being reflected in the kinds of genes being expressed there, which in turn is connected directly to the mechanisms that lead, through expression differences, to differential retention of pairs produced by different types of duplication (Tasdighian et al., 2017; Qiao et al., 2019).

Single Cell Data vs. Single Cell Type/State Data

Most of our results were based on root cell types and states, not on individual cells. This is a potential limitation, if it is desirable to assay gene expression at the most fundamental level

both qualitatively and quantitatively. The difficulties of doing so are both technical and biological. Technically, there is a tradeoff between number of cells and number of transcripts that can be assayed; in the system used to generate these data, the "dropout" effect is a particular problem, and can distort estimates of transcript numbers, particularly between highly and weakly expressed genes (Bhargava et al., 2014; Lähnemann et al., 2020).

Biologically, it is now known that, in contrast to early concepts of gene regulation, expression is stochastic and noisy, with bursts of transcription interspersed with inactivity, and the levels of transcripts, and thus proteins, depend on complex dynamics of expression and degradation (Araujo et al., 2017; Nicholson, 2019; Tunnacliffe and Chubb, 2020). This complexity has not been incorporated into models that seek to explain duplicate gene retention and loss. As an example, for a protein encoded by a pair of paralogs and for which dosage balance maintenance is important (e.g., as part of a multi-subunit complex), it is the sum of the expression of both paralogs that is critical; the numbers of transcripts produced from either paralog is less important in the short term, but potentially dictates whether one paralog will ultimately be lost in the process of fractionation (Gout and Lynch, 2015). This revised version of the duplication-degeneration-complementation (DDC) model (Force et al., 1999) involves gradual evolution of stochastic differences in paralog expression, but it is now clear that stochastic differences are also part of the transcriptional process itself, and that these also can play a role in promoting the preservation of duplicate genes (Rodrigo and Fares, 2018; Chapal et al., 2019).

A particular scA/scB transcript ratio for a paralog pair in a cell cluster could be due to all cells expressing that ratio (rheostat-like control of expression); to an on/off mechanism, with subsets of the cell population expressing scA and scB, creating the scA/scB ratio by modulating the relative number of cells expressing each gene (Nicholson, 2019); or to more complex modes of transcriptional control (Tunnacliffe and Chubb, 2020). For polyploids, particularly allopolyploids that combine diverged genomes, there exists the possibility that a single cell cluster might include individual cells that are differentiated in their expression by subgenome. This seems unlikely, given the integration observed for higher level phenotypes – an allopolyploid individual is not a mosaic for the two different floral morphologies of its progenitors, for example, but instead integrates the developmental gene networks of its progenitors to produce a distinctive flower. However, given the idea that duplicate gene regulatory networks can diverge from one another (Blanc and Wolfe, 2004), which we see here among root cell clusters (**Figure 8**), it is worth looking for evidence of higher-order independence of genome expression, and the level of the cell is a natural place to look. Despite our being able to analyze only a small number of gene pair × cell combinations with statistical rigor, it appears that although most gene pairs express both paralogs in individual cells (which is most consistent with a rheostat model), there do appear to be cases where expression of paralogs is partitioned into discrete cells (**Table 2**). As single cell technology improves, it will be interesting to look at larger numbers of cells for any evidence of differentiation by subgenome, and to quantify the extent of deterministic (i.e., rheostat) vs. stochastic (i.e., on/off) gene regulation by determining if differences in expression among cell types are achieved by changes in the per-cell expression level or by changes in the fraction of cells expressing the gene.

Conclusion and Future Prospects

The ability to assay transcript accumulation at the level of cell types and their developmental states permits elucidation of gene expression patterns at a very fine scale, and although the results we present here largely fall short of true single cell transcriptomics, it is almost certain that this and other current technical challenges will be overcome (Lähnemann et al., 2020). We can look forward soon to being able to compare the expression of gene pairs not only in root cells, but in all cells of the plant, eliminating the need to compare root single cell data with non-root transcriptomes generated from tissues or whole organs, as we did here.

Our characterization of the patterns of duplicate gene expression has identified a wealth of examples of differential expression of gene pairs among root cell types, the first step in achieving a detailed understanding of such issues as how pairs of homoeologs are regulated in allopolyploids at the level of trans-acting factors and cis-regulatory elements (Hu and Wendel, 2019). Single cell tools already exist for addressing these issues by assaying chromatin accessibility (Farmer et al., 2020) and mapping the epigenomic landscape of the cell (Zhou et al., 2019). For understanding functional differences between paralogs, and for even finer scale "sub-localization" of gene action within cells that may be a driver of paralog retention (Qiu et al., 2019), a host of other single cell-omics tools exist or are under development (Macaulay et al., 2017; Hasle et al., 2020).

We can learn much from single cell expression studies of a single accession of *A. thaliana*, but it is known that gene expression varies considerably among genotypes of this species (Cortijo et al., 2019), and that the response even to autopolyploidy in *Arabidopsis* varies among accessions (Yu et al., 2010; Song et al., 2020), so we look forward to the availability in expression atlases (e.g., Papatheodorou et al., 2019) of single cell data for roots and other tissues of additional individuals and accessions, and also from other species. Interspecific data would allow us to hypothesize ancestral states with far more confidence, particularly data from Cleomaceae, the sister family to Brassicaceae, whose common ancestor pre-dates the alpha WGD event (Edger et al., 2018).

Moreover, there are fascinating questions about the evolution of cell type transcriptomes that can only be addressed with comparative studies. For example, Liang et al. (2018) describe the phenomenon of "correlated evolution" of transcriptomes, in which transcriptomes of different (non-homologous) tissues of a species cluster together, instead of transcriptomes of homologous tissues in different species clustering together as expected. This is thought to be a consequence of the non-independence of transcriptomes in cell types that share transcription factors and their target genes. It is reminiscent of concerted evolution in gene families and concerted divergence of gene regulatory networks, both of which are phenomena of considerable interest and importance in polyploids (Blanc and Wolfe, 2004; Qiao et al., 2019). Comparative single cell data from other *Arabidopsis* species, including from recently formed

allopolyploids in this excellent model system, hold much promise for addressing these and other questions.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors conceived and wrote the manuscript. JC and AF conducted most analyses.

FUNDING

JS acknowledges support from National Science Foundation grants IOS-1444400 and IOS-1923589. JD and JC's work on polyploidy has been supported by the United States National Science Foundation, most recently by award 1257522.

ACKNOWLEDGMENTS

We thank Marc Libault for critical reading of the manuscript and for inspiring this project. We are grateful to a number of colleagues who provided advice, answered questions, or provided unpublished materials: Jonathan Wendel, Andy Paterson, James Schnable, Pat Edger, Nicholas Panchy, and Shin-Han Shiu. We also thank two reviewers for their helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.596150/full#supplementary-material

Supplementary Figure 1 | Correlation between fractions of genes expressed per RCC calculated using two different thresholds of expression (≥1 UMI and >1 cell).

Supplementary Figure 2 | Correlation between genes expressed per RCC and cells per RCC.

Supplementary Figure 3 | Distributions of root cell clusters expressed per gene **(A)** or gene pair **(B)** for each duplication mechanism.

Supplementary Figure 4 | Bulk tissue RNA-seq based estimates of expression for Class 1 genes. **(A)** Expression breadth (number of SRA libraries out of the 214 libraries examined expressing the gene at ≥5 reads). **(B)** Expression breadth in non-root tissues (number of SRA libraries out of the 92 non-root tissue containing SRA libraries examined expressing the gene at ≥5 reads). **(C)** Expression level (combined read count from all 214 SRA libraries). **(D)** Expression level in non-root tissues (combined read count from 92 non-root tissue containing SRA libraries). "Class_1_scA" are the copies of Class 1 gene pairs expressed in the single cell root RCCs. "Class_1_scB" are the paralogs of Class 1 gene pairs not expressed in root RCCs. "Other" is the set of all other nuclear genes in *Arabidopsis* (Araport 11).

Supplementary Figure 5 | RCC-u genes are enriched for WGD duplicates. Blue bars and orange bars indicate the fraction of genes that retain duplicates from the indicated mechanism for the whole genome and the RCC-u genes, respectively. The RCC-u gene set has a higher fraction of WGD duplicates and older TEs, and lower fraction of SSDs and singletons, relative to the whole genome.

Supplementary Table 1 | Single cell clustering into the 36 RCCs.

Supplementary Table 2 | Bulk tissue expression data.

Supplementary Table 3 | Single cell expression data, properties of duplicate gene pairs, and GO enrichment data for Class 4 pairs.

Supplementary Table 4 | Counts of paralogous gene pairs by expression class and duplication type.

Supplementary Table 5 | Counts of scB expression for Class I paralogs.

Supplementary Table 6 | K_a/K_s by expression class and duplication type.

Supplementary Table 7 | Paralog expression patterns across developmental states within cell types.

Supplementary Table 8 | Breakdown by expression class and duplication mechanism of 4,599 Wang et al. (2013) paralog pairs ubiquitously expressed in root cell clusters (RCC-u gene pairs).

Supplementary Table 9 | Frequency of expression bias by duplication type for RCC-u and non-RCC-u pairs.

REFERENCES

Adams, K. L., Cronn, R., Percifield, R., and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4649–4654. doi: 10.1073/pnas.0630618100

Araujo, I. S., Pietsch, J. M., Keizer, E. M., Greese, B., Balkunde, R., Fleck, C., et al. (2017). Stochastic gene expression in *Arabidopsis thaliana*. *Nat. Commun.* 8:2132.

Arendt, D., Musser, J. M., Baker, C. V., Bergman, A., Cepko, C., Erwin, D. H., et al. (2016). The origin and evolution of cell types. *Nat. Rev. Genet.* 17, 744–757.

Bhargava, V., Head, S. R., Ordoukhanian, P., Mercola, M., and Subramaniam, S. (2014). Technical variations in low-input RNA-seq methodologies. *Sci. Rep.* 4:3678.

Birchler, J. A., and Veitia, R. A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 186, 54–62. doi: 10.1111/j.1469-8137.2009.03087.x

Birchler, J. A., and Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14746–14753. doi: 10.1073/pnas.1207726109

Birchler, J. A., and Veitia, R. A. (2014). The Gene Balance Hypothesis: dosage effects in plants. *Methods Mol. Biol.* 1112, 25–32. doi: 10.1007/978-1-62703-773-0_2

Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., Galbraith, D. W., et al. (2003). A gene expression map of the *Arabidopsis* root. *Science* 302, 1956–1960. doi: 10.1126/science.1090022

Blanc, G., and Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16, 1679–1691. doi: 10.1105/tpc.021410

Buggs, R. J., Wendel, J. F., Doyle, J. J., Soltis, D. E., Soltis, P. S., and Coate, J. E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* 369:20130354. doi: 10.1098/rstb.2013.0354

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. doi: 10.1038/nbt.4096

Chapal, M., Mintzer, S., Brodsky, S., Carmi, M., and Barkai, N. (2019). Resolving noise-control conflict by gene duplication. *PLoS Biol.* 17:e3000289. doi: 10.1371/journal.pbio.3000289

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tpj.13415

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4, 258–268. doi: 10.1038/s41477-018-0136-7

Coate, J. E., Song, M. J., Bombarely, A., and Doyle, J. J. (2016). Expression-level support for gene dosage sensitivity in three *Glycine* subgenus *Glycine* polyploids and their diploid progenitors. *New Phytol.* 212, 1083–1093. doi: 10.1111/nph.14090

Cortijo, S., Aydin, Z., Ahnert, S., and Locke, J. C. (2019). Widespread inter-individual gene expression variability in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 15:e8591.

Defoort, J., Van de Peer, Y., and Carretero-Paulet, L. (2019). The Evolution of Gene Duplicates in Angiosperms and the Impact of Protein-Protein Interactions and the Mechanism of Duplication. *Genome Biol. Evol.* 11, 2292–2305.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M. C. P. (2019). Spatiotemporal Developmental Trajectories in the *Arabidopsis* Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev. Cell.* 48, 840-852.e845.

Edger, P. P., Hall, J. C., Harkess, A., Tang, M., Coombs, J., Mohammadin, S., et al. (2018). Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am. J. Bot.* 105, 463–469. doi: 10.1002/ajb2.1040

Efroni, I., and Birnbaum, K. D. (2016). The potential of single-cell profiling in plants. *Genome Biol.* 17, 1–8. doi: 10.1002/9783527678679.dg11957

Emery, M., Willis, M. M. S., Hao, Y., Barry, K., Oakgrove, K., Peng, Y., et al. (2018). Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genet.* 14:e1007267. doi: 10.1371/journal.pgen.1007267

Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, J., and Libault, M. (2020). The impact of chromatin remodeling on gene expression at the single cell level in *Arabidopsis thaliana*. *bioRxiv [Preprint]* doi: 10.1101/2020.07.27.223156

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Gabaldon, T., and Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14, 360–366. doi: 10.1038/nrg3456

Gallagher, J. P., Grover, C. E., Hu, G., Jareczek, J. J., and Wendel, J. F. (2020). Conservation and divergence in duplicated fiber coexpression networks accompanying domestication of the polyploid *Gossypium hirsutum* L. *G3* 10, 2879–2892. doi: 10.1534/g3.120.401362

Gou, J. Y., Wang, L. J., Chen, S. P., Hu, W. L., and Chen, X. Y. (2007). Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res.* 17, 422–434. doi: 10.1038/sj.cr.7310150

Gout, J. F., and Lynch, M. (2015). Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol. Biol. Evol.* 32, 2141–2148. doi: 10.1093/molbev/msv095

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Hanada, K., Kuromori, T., Myouga, F., Toyoda, T., and Shinozaki, K. (2009). Increased expression and protein divergence in duplicate genes is associated with morphological diversification. *PLoS Genet.* 5:e1000781. doi: 10.1371/journal.pgen.1000781

Hao, Y., Washburn, J. D., Rosenthal, J., Nielsen, B., Lyons, E., Edger, P. P., et al. (2018). Patterns of population variation in two paleopolyploid eudicot lineages suggest that dosage-based selection on homeologs is long-lived. *Genome Biol. Evol.* 10, 999–1011. doi: 10.1093/gbe/evy061

Hasle, N., Cooke, A., Srivatsan, S., Huang, H., Stephany, J. J., Krieger, Z., et al. (2020). High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol. Syst. Biol.* 16:e9442.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* 5:13.

Hossain, M. S., Joshi, T., and Stacey, G. (2015). System approaches to study root hairs as a single cell plant model: current status and future perspectives. *Front. Plant Sci.* 6:363. doi: 10.3389/fpls.2015.00363

Hovav, R., Udall, J. A., Chaudhary, B., Rapp, R., Flagel, L., and Wendel, J. F. (2008a). Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6191–6195. doi: 10.1073/pnas.0711569105

Hovav, R., Udall, J. A., Hovav, E., Rapp, R., Flagel, L., and Wendel, J. F. (2008b). A majority of cotton genes are expressed in single-celled fiber. *Planta* 227, 319–329. doi: 10.1007/s00425-007-0619-7

Hu, G., and Wendel, J. F. (2019). Cis-trans controls and regulatory novelty accompanying allopolyploidization. *New Phytol.* 221, 1691–1700. doi: 10.1111/nph.15515

Huang, L., and Schiefelbein, J. (2015). Conserved Gene Expression Programs in Developing Roots from Diverse Plants. *Plant Cell* 27, 2119–2132. doi: 10.1105/tpc.15.00328

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689

Jean-Baptiste, K., McFaline-Figueroa, J. L., Alexandre, C. M., Dorrity, M. W., Saunders, L., Bubb, K. L., et al. (2019). Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* 31, 993–1011. doi: 10.1105/tpc.18.00785

Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J. E., McKain, M. R., McNeal, J., et al. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* 13:R3.

Kenrick, P., and Strullu-Derrien, C. (2014). The origin and early evolution of roots. *Plant Physiol.* 166, 570–580. doi: 10.1104/pp.114.244517

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20:278.

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21:31.

Langham, R. J., Walsh, J., Dunn, M., Ko, C., Goff, S. A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics* 166, 935–945. doi: 10.1534/genetics.166.2.935

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Leinonen, R., Sugawara, H., and Shumway, M. (2011). International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019

Liang, C., Musser, J. M., Cloutier, A., Prum, R. O., and Wagner, G. P. (2018). Pervasive correlated evolution in gene expression shapes cell and tissue type transcriptomes. *Genome Biol. Evol.* 10, 538–552. doi: 10.1093/gbe/evy016

Liang, Z., and Schnable, J. C. (2018). Functional divergence between subgenomes and gene pairs after whole genome duplications. *Mol. Plant* 11, 388–397. doi: 10.1016/j.molp.2017.12.010

Libault, M., Pingault, L., Zogli, P., and Schiefelbein, J. (2017). Plant systems biology at the single-cell level. *Trends Plant Sci.* 22, 949–960. doi: 10.1016/j.tplants.2017.08.006

Lloyd, J. P., Tsai, Z. T., Sowers, R. P., Panchy, N. L., and Shiu, S. H. (2018). A Model-Based Approach for Identifying Functional Intergenic Transcribed Regions and Noncoding RNAs. *Mol. Biol. Evol.* 35, 1422–1436. doi: 10.1093/molbev/msy035

Luecken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15:e8746.

Lynch, M. (2020). The evolutionary scaling of cellular traits imposed by the drift barrier. *Proc. Natl. Acad. Sci. U.S.A.* 117, 10435–10444. doi: 10.1073/pnas.2000446117

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* 159, 1789–1804.

Lynch, M., and Trickovic, B. (2020). A theoretical framework for evolutionary cell biology. *J. Mol. Biol.* 432, 1861–1879. doi: 10.1016/j.jmb.2020.02.006

Macaulay, I. C., Ponting, C. P., and Voet, T. (2017). Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33, 155–168. doi: 10.1016/j.tig.2016.12.003

McClintock, B. (1984). The Significance of Responses of the Genome to Challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

Morris, S. A. (2019). The evolving concept of cell identity in the single cell era. *Development* 146:dev169748. doi: 10.1242/dev.169748

Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.

Nicholson, D. J. (2019). Is the cell really a machine? *J. Theor. Biol.* 477, 108–126. doi: 10.1016/j.jtbi.2019.06.002

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer.

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316.

Panchy, N. L., Azodi, C. B., Winship, E. F., O'Malley, R. C., and Shiu, S. H. (2019). Expression and regulatory asymmetry of retained *Arabidopsis thaliana* transcription factor genes derived from whole genome duplication. *BMC Evol. Biol.* 19:77. doi: 10.1186/s12862-019-1398-z

Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M.-P., George, N., Fexova, S., et al. (2019). Expression Atlas update: from tissues to single cells. *Nucleic Acids Res.* 48, D77–D83.

Papp, B., Pál, C., and Hurst, L. D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197. doi: 10.1038/nature01771

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* 20:38.

Qiao, Z., and Libault, M. (2013). Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. *Front. Plant Sci.* 4:484. doi: 10.3389/fpls.2013.00484

Qiu, Y., Tay, Y. V., Ruan, Y., and Adams, K. L. (2019). Divergence of duplicated genes by repeated partitioning of splice forms and subcellular localization. *New Phytol.* 225, 1011–1022. doi: 10.1111/nph.16148

Raven, J. A., and Edwards, D. (2001). Roots: evolutionary origins and biogeochemical significance. *J. Exp. Bot.* 52, 381–401. doi: 10.1093/jxb/52.suppl_1.381

Renny-Byfield, S., Gallagher, J. P., Grover, C. E., Szadkowski, E., Page, J. T., Udall, J. A., et al. (2014). Ancient gene duplicates in Gossypium (cotton) exhibit near-complete expression divergence. *Genome Biol. Evol.* 6, 559–571. doi: 10.1093/gbe/evu037

Rodrigo, G., and Fares, M. A. (2018). Intrinsic adaptive value and early fate of gene duplication revealed by a bottom-up approach. *eLife* 7:e29739.

Ryu, K. H., Huang, L., Kang, H. M., and Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* 179, 1444–1456. doi: 10.1104/pp.18.01482

Schnable, J. C., Wang, X., Pires, J. C., and Freeling, M. (2012). Escape from preferential retention following repeated whole genome duplications in plants. *Front. Plant Sci.* 3:94. doi: 10.3389/fpls.2012.00094

Shen, L., and Sinai, M. (2019). *GeneOverlap: Test and Visualize Gene Overlaps. R Package Version 1.23.0*. Available at: http://shenlab-sinai.github.io/shenlab-sinai/. https://doi.org/doi:10.18129/B9.bioc.GeneOverlap (accessed June 3, 2020).

Shi, Y. H., Zhu, S. W., Mao, X. Z., Feng, J. X., Qin, Y. M., Zhang, L., et al. (2006). Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell.* 18, 651–664. doi: 10.1105/tpc.105.040303

Shulse, C. N., Cole, B. J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., et al. (2019). High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* 27, 2241-2247.e2244.

Song, M. J., Potter, B. I., Doyle, J. J., and Coate, J. E. (2020). Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. *Plant Cell* 32, 1434–1448. doi: 10.1105/tpc.19.00832

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 12, 1611–1618. doi: 10.1101/gr.361602

Steige, K. A., and Slotte, T. (2016). Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. *Curr. Opin. Plant Biol.* 30, 88–93. doi: 10.1016/j.pbi.2016.02.006

Taliercio, E. W., and Boykin, D. (2007). Analysis of gene expression in cotton fiber initials. *BMC Plant Biol.* 7:22. doi: 10.1186/1471-2229-7-22

Tasdighian, S., Van Bel, M., Li, Z., Van de Peer, Y., Carretero-Paulet, L., and Maere, S. (2017). Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell* 29, 2766–2785. doi: 10.1105/tpc.17.00313

Tunnacliffe, E., and Chubb, J. R. (2020). What is a transcriptional burst? *Trends Genet.* 36, 288–297. doi: 10.1016/j.tig.2020.01.003

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

Vickaryous, M. K., and Hall, B. K. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev. Cambridge Philos. Soc.* 81, 425–455. doi: 10.1017/s1464793106007068

Wang, Y., Tan, X., and Paterson, A. H. (2013). Different patterns of gene structure divergence following gene duplication in Arabidopsis. *BMC Genomics* 14:652. doi: 10.1186/1471-2164-14-652

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* 22, 209–212. doi: 10.1080/01621459.1927.10502953

Woodhouse, M. R., Cheng, F., Pires, J. C., Lisch, D., Freeling, M., and Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl. Acad. Sci. U.S.A.* 111:201402475.

Xie, J., Li, Y., Liu, X., Zhao, Y., Li, B., Ingvarsson, P. K., et al. (2019). Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. *Plant Cell* 31, 563–578. doi: 10.1105/tpc.18.00601

Yoo, M. J., Szadkowski, E., and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

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

Yu, Z., Haberer, G., Matthes, M., Rattei, T., Mayer, K. F., Gierl, A., et al. (2010). Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17809–17814.

Yuan, M., Yang, X., Lin, J., Cao, X., Chen, F., Zhang, X., et al. (2020). Alignment of cell lineage trees elucidates genetic programs for the development and evolution of cell types. *iScience* 23, 101273–101273. doi: 10.1016/j.isci.2020.101273

Zhang, T. Q., Xu, Z. G., Shang, G. D., and Wang, J. W. (2019). A Single-Cell RNA sequencing profiles the developmental landscape of arabidopsis root. *Mol. Plant* 12, 648–660. doi: 10.1016/j.molp.2019.04.004

Zhou, S., Jiang, W., Zhao, Y., and Zhou, D. X. (2019). Single-cell three-dimensional genome structures of rice gametes and unicellular zygotes. *Nat. Plants* 5, 795–800. doi: 10.1038/s41477-019-0471-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 November 2020
doi: 10.3389/fgene.2020.564928

Inbreeding Depression in Genotypically Matched Diploid and Tetraploid Maize

*Hong Yao[1], Sanvesh Srivastava[2], Nathan Swyers[1], Fangpu Han[1], Rebecca W. Doerge[3] and James A. Birchler[1]**

[1] Division of Biological Sciences, University of Missouri, Columbia, MO, United States, [2] Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA, United States, [3] Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, United States

The genetic and molecular basis of heterosis has long been studied but without a consensus about mechanism. The opposite effect, inbreeding depression, results from repeated self-pollination and leads to a reduction in vigor. A popular explanation for this reaction is the homozygosis of recessive, slightly deleterious alleles upon inbreeding. However, extensive studies in alfalfa indicated that inbreeding between diploids and autotetraploids was similar despite the fact that homozygosis of alleles would be dramatically different. The availability of tetraploid lines of maize generated directly from various inbred lines provided the opportunity to examine this issue in detail in perfectly matched diploid and tetraploid hybrids and their parallel inbreeding regimes. Identical hybrids at the diploid and tetraploid levels were inbred in triplicate for seven generations. At the conclusion of this regime, F1 hybrids and selected representative generations (S1, S3, S5, S7) were characterized phenotypically in randomized blocks during the same field conditions. Quantitative measures of the multiple generations of inbreeding provided little evidence for a distinction in the decline of vigor between the diploids and the tetraploids. The results suggest that the homozygosis of completely recessive, slightly deleterious alleles is an inadequate hypothesis to explain inbreeding depression in general.

Keywords: maize, tetraploid, heterosis, progressive heterosis, inbreeding depression

OPEN ACCESS

Edited by:
Yves Van de Peer,
Ghent University, Belgium

Reviewed by:
Charlie Brummer,
University of California, Davis, United States
Jianming Yu,
Iowa State University, United States

***Correspondence:**
James A. Birchler
BirchlerJ@Missouri.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *22 May 2020*
Accepted: *13 October 2020*
Published: *30 November 2020*

Citation:
Yao H, Srivastava S, Swyers N, Han F, Doerge RW and Birchler JA (2020) Inbreeding Depression in Genotypically Matched Diploid and Tetraploid Maize. Front. Genet. 11:564928. doi: 10.3389/fgene.2020.564928

INTRODUCTION

Heterosis refers to the phenomenon that hybrid progeny of inbred parents will exceed the performance of the better parent (Shull, 1908; Bruce, 1910; Chen, 2013). It has been capitalized upon by plant breeders for decades to enhance yields, but its genetic and molecular basis has escaped understanding. A popular concept to explain heterosis has been that slightly deleterious homozygous mutations that differ in the parents are complemented in the hybrid (Jones, 1917). To the degree that deleterious mutations are present and homozygous in the different inbreds, this complementation will certainly occur. However, modern inbreds might have been purged of the obviously deleterious mutations from heterotic groups but still exhibit a robust heterotic effect when crossing inbreds from the opposite groups (Duvick, 1999). This concept is derived from work on heterosis in ostensibly diploid species.

Indeed, the behavior of heterosis in polyploids is not prima facie explicable on this hypothesis (East, 1936; Briggle, 1963; Li et al., 2008). First, the phenomenon of progressive heterosis has been

documented in several tetraploid plants including alfalfa, potato, and maize (Levings et al., 1967; Dunbier and Bingham, 1975; Mok and Peloquin, 1975; Bingham, 1980; Chase, 1980; Groose et al., 1989; Sockness and Dudley, 1989a,b; Bingham et al., 1994; Riddle et al., 2010; Washburn et al., 2019). This phenomenon is that double-cross hybrids resulting from a cross of two different single cross hybrids exhibit a further increase in heterosis. Symbolically, this can be illustrated in that an ABCD hybrid shows greater heterosis than AABB or CCDD. In contrast, double-cross hybrids at the diploid level do not routinely show this response (Washburn et al., 2019). In order for complementation to explain progressive heterosis, the AAAA homozygous lines would need to have different detrimental alleles than BBBB but some in common. At the same time CCCC would need to have a different set than DDDD but some different alleles in common so that AABB and CCDD would, respectively, still be homozygous for different sets to complement in the double-cross hybrid. All these conditions would need to be met in order not to reconstitute a homozygous state for detrimental recessives (Washburn et al., 2019).

A second observation that is not explained by the complementation concept is that heterosis is different in reciprocal triploid hybrids despite being quite similar in diploid hybrids (i.e., AAB $\neq$ BBA but AB = BA) (Yao et al., 2013; Tan et al., 2016). If recessive detrimentals were the sole basis of heterosis, then AAB and BBA should be similar. The fact that they are routinely distinct indicates that there is a dosage component to heterosis.

The third observation of note from polyploidy heterosis is that inbreeding depression curves of matched diploid and tetraploid genotypes are quite similar despite a very different predicted rate of homozygosis. This effect has been studied in alfalfa, wheatgrass, and to a lesser degree in maize (Alexander and Sonnemaker, 1961; Demarly, 1963; Busbice and Wilsie, 1966; Dewey, 1966, 1969; Rice and Dudley, 1974; Chase, 1980; Gallais, 1984; Li and Brummer, 2009). Indeed, in the early days of alfalfa breeding, this parallel led to great confusion (Williams, 1931; Tysdal et al., 1942). For a single gene in a diploid, self-pollination of a heterozygote would produce a quarter of the progeny that are homozygous for either of the two alleles present (Charlesworth and Willis, 2009). However, in a autotetraploid duplex hybrid (AABB), self-pollination would produce only 1/36 of the progeny that would be homozygous for either allele for genes near the centromeres. This tetraploid estimate is subject to many caveats such as recombination between a locus being followed and the centromere as well as the mode of segregation of the four homologous chromosomes during meiosis I. Nevertheless, during an inbreeding regime the rate of homozygosis in an autotetraploid would be predicted to be considerably slower than in a diploid (Welch, 1962; Levings and Alexander, 1966; Doyle, 1979, 1986; Birchler, 2012). The fact that the decline in vigor between diploids and autotetraploids is quite similar suggests an explanation for heterosis and inbreeding depression needs further explanation than complementation of recessive mutations (Washburn et al., 2019).

The subject of the present study was a test of the rate of inbreeding depression between diploids and autotetraploids that were directly derived from the diploid inbreds and thus of exactly the same genotype. Previous studies in maize were of a preliminary nature and the tetraploids might have some differences with the diploids with which they were compared. The results of the present study reveal that indeed there is a very similar rate of inbreeding depression upon selfing of comparable genotypes of the starting hybrid materials at the diploid and tetraploid levels.

MATERIALS AND METHODS

The experiments to investigate inbreeding depression rates in diploid and tetraploid maize lines were conducted in 2009 (**Supplementary Material 1**) and 2008 (**Supplementary Material 2**) in Columbia, Missouri. The four diploid and tetraploid parental inbred maize lines A188 (2x, 4x), Oh43 (2x, 4x), B73 (2x, 4x), and W22 (2x, 4x) were used in this study (Kato and Birchler, 2006; Riddle et al., 2006; Riddle and Birchler, 2008). The following F1 hybrids from these parental lines were grown: Oh43/A188 (2x), A188/Oh43 (2x), W22/B73 (2x), B73/W22 (2x), Oh43/W22 (2x), W22/Oh43 (2x), W22/A188 (2x), A188/W22 (2x), B73/A188 (2x), A188/B73 (2x), B73/Oh43 (2x), Oh43/B73 (2x), A188/Oh43 (2x) $\times$ B73/W22 (2x), B73/W22 (2x) $\times$ A188/Oh43 (2x), W22/B73 (4x), A188/Oh43 (4x), Oh43/A188/W22/B73 (4x).

Each F1 hybrid line was previously self-mated for seven generations and progenies from generations 1, 3, 5, and 7 (named S1, S3, S5, and S7) were used for data collection. Genetic segregation occurred after the first generation of the self-mating population in this experiment. Kernels from three different S2 ears (resulting from self-mating S1 plants) were used to produce the S3–S7 lines, to account for the genetic diversity among the S1 plants derived from the same F1 hybrid. Thus, there were three selfing lineages for each genotype.

The experiments were based on a randomized complete block design. The maize lines from all ploidy levels, genotypes, and generations were planted in three fields. Each maize line was grown in each of the three fields (blocks) with border rows of unrelated maize. All genotypes were randomized within the blocks with intermixing of the diploid and tetraploid samples. Soil type was "Leonard silt loam" or "Mexico silt loam." Twenty seeds of the respective maize lines were planted per row with 22.96 cm spacing between plants in a row, and, whenever possible, data from at most 12 plants were collected. The planting dates of the three blocks in 2009 were May 21, June 1, and June 14 at the University of Missouri Genetic Farm near Columbia, Missouri.

Data on the following phenotypes were collected (the names in parentheses denote the names used in the analysis for the corresponding phenotype): (1) the number of days to anther emergence after planting (flowering time); (2) the number of days to silk emergence after planting (silk emergence time); (3) the ear length of the maize plant (ear length); (4) the tassel branch number (tassel branch number); (5) the height of the plant to the growing tip at the fourth week (4th week height); (6) the height of the plant to the growing tip at the sixth week (6th week height);

(7) the height of the adult plant to the top of the tassel (adult plant height); (8) the length of the fifth leaf from the top of the adult plant (length of the 5th leaf from the top), and (9) the width of the fifth leaf from the top of the adult plant (width of the 5th leaf from the top).

The experiments were conducted to investigate the following biological questions about inbreeding depression rates:

1. Is the inbreeding depression different between diploid and tetraploid lines with the same genetic constitution?
2. Is the inbreeding depression different between lines with different genetic constitution but the same ploidy?
3. Does inbreeding depression occur in all the measured phenotypes?
4. Is there inbreeding depression in every diploid and tetraploid genotype?
5. How is the inbreeding depression rate affected by ploidy, genetic constitution, and the interaction between ploidy and genetic constitution?
6. Are there any parental effects on inbreeding depression rate?
7. Are the S7 lines different from their corresponding progenitor inbred lines?

Statistical Analyses

We have summarized the data for every field and phenotype by averaging over the biological replicates. Separately, the summarized data from the different fields and years are analyzed. The generations F0, S1, S3, S5, and S7 are labeled as 0, 1, 3, 5, and 7, respectively. Let y be the observed phenotypic value in generation $gen \in \{0, 1, 3, 5, 7\}$ and *ploidy* be an indicator variable that is 0 and 1 for diploid and tetraploid plants, respectively. If ϵ and *geno* are the idiosyncratic error in measuring the phenotype and dummy variable denoting the genotype of the plant, then we model y using two different models depending on the question:

$$y = \beta_0 + \beta_1 gen + \beta_2 ploidy + \beta_3 \left(gen \times ploidy\right) + \varepsilon,$$

$$y = \beta_0 + \beta_1 gen + \beta_2 geno + \beta_3 \left(gen \times geno\right) + \varepsilon,$$

where β_1, β_2, β_3 are the regression coefficients; β_0 is the intercept; and the former and latter models account for the interaction of generation with ploidy and generation with genotype, respectively. If inbreeding depression is present in a phenotype, then β_1 is negative. Our hypotheses tests are performed under the additional assumption that ϵ is Gaussian with mean 0. In both models, our questions are answered by testing the null hypothesis $\beta_3 = 0$; see **Supplementary Materials 1–4** for greater details.

Segregation Patterns

In order to test whether the chromosome segregation from the tetraploid hybrids followed tetrasomic or disomic patterns, progeny from self-pollinated ears of a W22/B73 hybrid were screened for the distribution of W22 or B73 chromosomes using fluorescence *in situ* hybridization (FISH). The following is a description of the distinctions between the two inbred lines for chromosomes that differ in cytological features (see also Albert et al., 2010).

Chromosome 1: Chromosome 1 can be distinguished between W22 and B73 by the strong TAG probe signal on the long arm of W22 chromosome 1 as compared to B73 chromosome 1 as well as the 4-12-1 probe signal on the short arm of W22 chromosome 1.

Chromosome 2: Chromosome 2 can be distinguished by the presence of the 5S gene cluster and then the strong CentC probe signal on W22 chromosome 2 contrasted by the very weak CentC signal on B73 chromosome 2 distinguishing the two genotypes from each other.

Chromosome 4: Chromosome 4 can be distinguished by the presence of a unique centromeric sequence (Cent4) and then the distinctive knob highlighted by DAPI stain on W22 chromosome 4 as well as the strong microsatellite TAG probe signal on the short arm of B73 chromosome 4 allowing the two genotypes to be determined.

Chromosome 5: B73 chromosome 5 displays a much stronger 4-12-1 probe signal on its short arms than its W22 counterpart. There also is a much stronger knob signal revealed by DAPI staining on W22 chromosome 5 than B73 chromosome 5.

Chromosome 8: W22 chromosome 8 has a much stronger CentC probe signal than B73 chromosome 8. Chromosome 8 can be determined by a consistent subtelomeric signal on the long arm.

Chromosome 9: Chromosome 9 has a consistent terminal knob on the short arm, but W22 chromosome 9 has a much stronger CentC probe signal than B73 chromosome 9.

Chromosome 10: Chromosome 10 is the shortest chromosome and can be distinguished from the other chromosomes by the absence of any consistent markers. W22 chromosome 10 has a much stronger CentC probe signal than B73 chromosome 10.

Fluorescence *in situ* Hybridization (FISH)

FISH was conducted as described (Kato et al., 2004). The set of probes used in this experiment is as follows: The CentC probe was labeled green and identifies a centromeric repeat. The 4-12-1 probe is labeled green and consists of a subtelomeric repeat. The NOR probe is labeled green and denotes the nucleolar organizing region. The TAG satellite probe is labeled red. The Cent4 probe is labeled red and shows a centromeric repeat specific to chromosome 4. Finally, DAPI is used to stain heterochromatin. Not all of the chromosome pairs could be accurately distinguished between W22 and B73, so those were not included in the analysis. Chromosomes 1, 2, 4, 5, 8, 9, and 10 were used because parental genotypes were able to be distinguished in the tetraploid. Green probes are labeled using AlexaFluor dUTPs, and red probes are labeled using TexasRed dCTPs.

RESULTS

As detailed in **Supplementary Materials 1–4**, the null hypothesis that the rate of inbreeding depression between diploids and tetraploids is different could be rejected in the vast majority of comparisons of genotypes in the four generations of selfing

progression that were analyzed (S1, S3, S5, S7). The different genotypes, however, could be distinguished from each other at the diploid level, but some comparisons at the tetraploid level were not significant (see **Supplementary Materials 1–3**. **Supplementary Material 4** contains the probabilities in tabular form.). All of the measured phenotypic characteristics exhibited inbreeding depression and the depression was observed at both ploidy levels examined. There was not sufficient data to make a generalization of whether the S7 was different from the corresponding inbred lines.

Figure 1 illustrates one comparison of adult plant height from the F1 to the S7 for the W22/B73 diploid and tetraploid, the

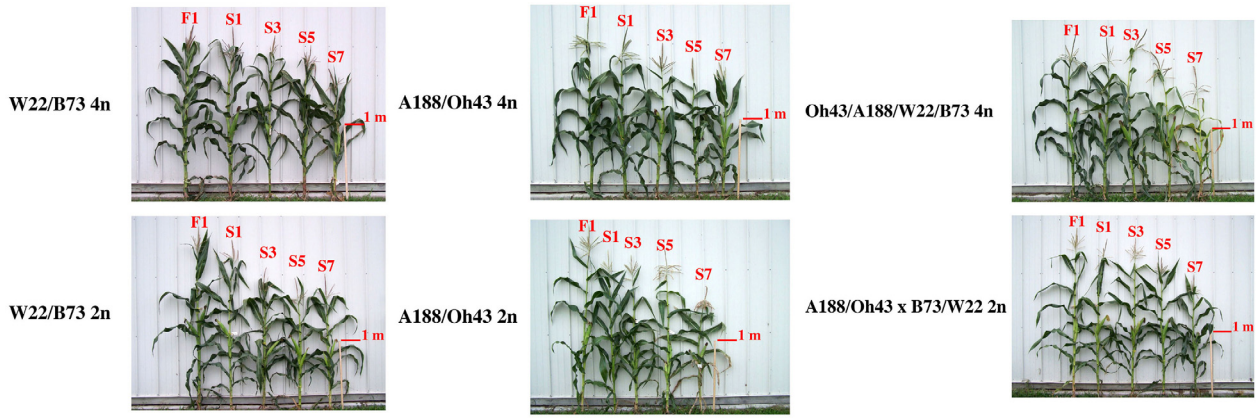


FIGURE 1 | Comparison of inbreeding of a W22/B73 hybrid for seven generations at the tetraploid and diploid levels. (Top) Examples of plants from the inbred generations from the F1 tetraploid to the seventh selfed generation. (Bottom) Examples from the diploid progression. One meter stick provides scale.

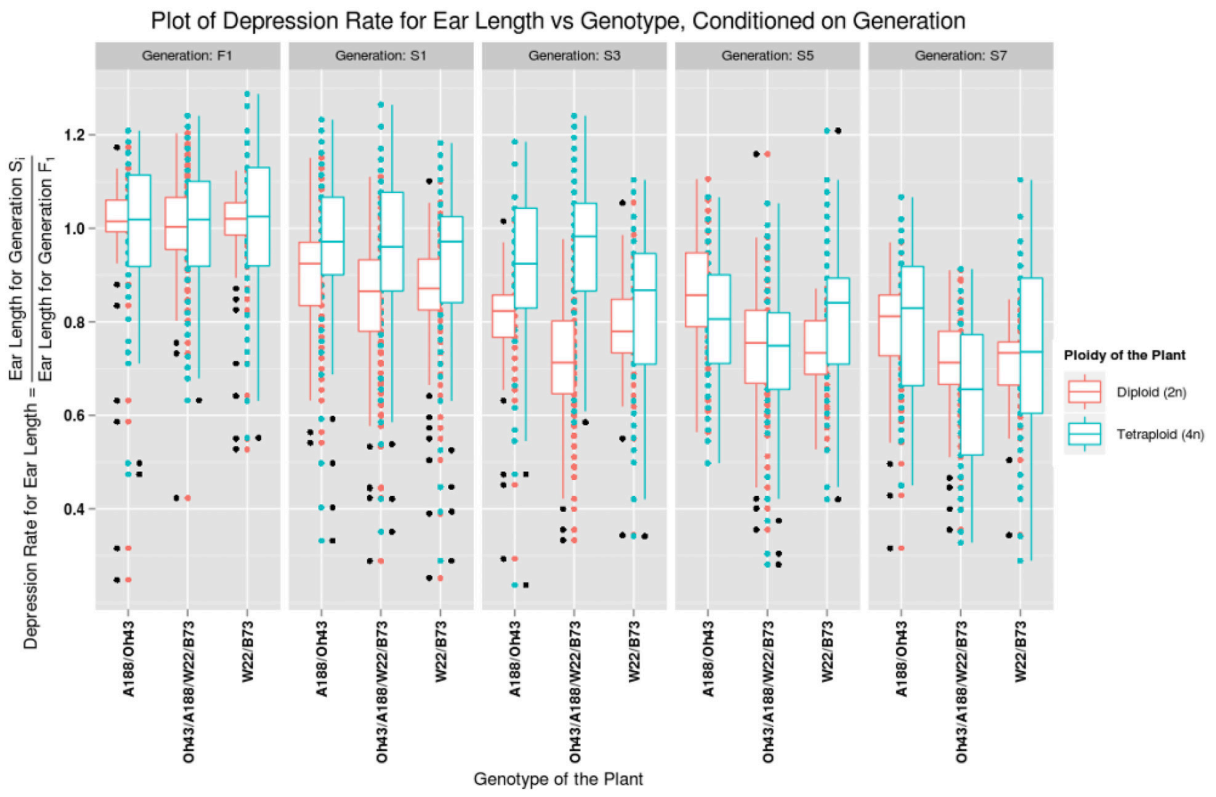


FIGURE 2 | Plot of depression rate for ear length vs. the genotype of the plant, conditioned on generations F1, S1, S3, S5, and S7. The data for diploid plants are graphed in red and the tetraploid plants are graphed in blue. The patterns for a particular ploidy can be observed by looking at the box plots of the corresponding color. The generation varies from F1 to S7 column-wise.

A188/Oh43 diploid and tetraploid, and the double-cross hybrid at the diploid and tetraploid levels. **Figure 2** shows the matched diploid and tetraploid measurements for the three comparisons from the F1 generation to the S7 for ear length. The plots of depression rate for the other characteristics measured are in **Supplementary Materials 1–3**.

In further analysis, the averaged phenotypic values across biological replicates were plotted across generations for the various traits examined [**Figure 3** (2008 data) and **Figure 4** (2009 data); **Supplementary Materials 3, 4**] separated into the three genotypes. The linear regression lines for the two ploidies are not significantly different for most traits and genotypes between diploid and tetraploid. A potential exception is the trait of the tassel branch number. However, the slope of the tetraploid is steeper than that of the diploid in those cases in which they differ suggesting a stronger effect of inbreeding in the tetraploids

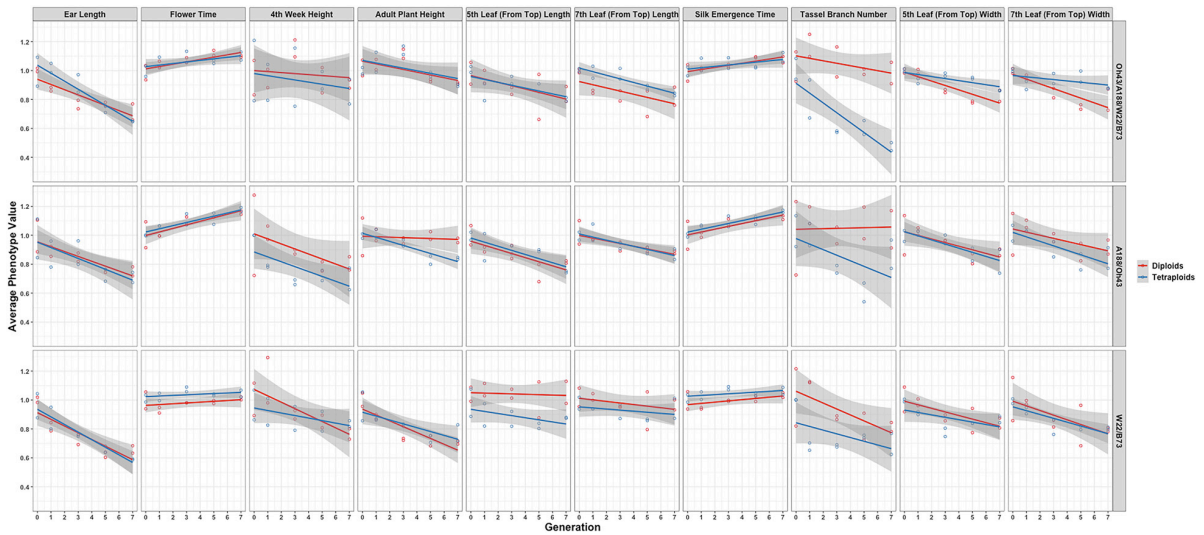


FIGURE 3 | Visualization of the interaction between ploidy and genotype in the 2008 data. Generation (*x*-axis) vs. averaged phenotypic values across biological replicates (*y*-axis) conditioned on the genetic constitution (rows) and trait (columns). Two linear regression lines are superimposed on every panel and their color indicates the diploid (red) and tetraploid plants (blue). The gray-colored band on a regression indicates a 95% confidence interval for the whole line.

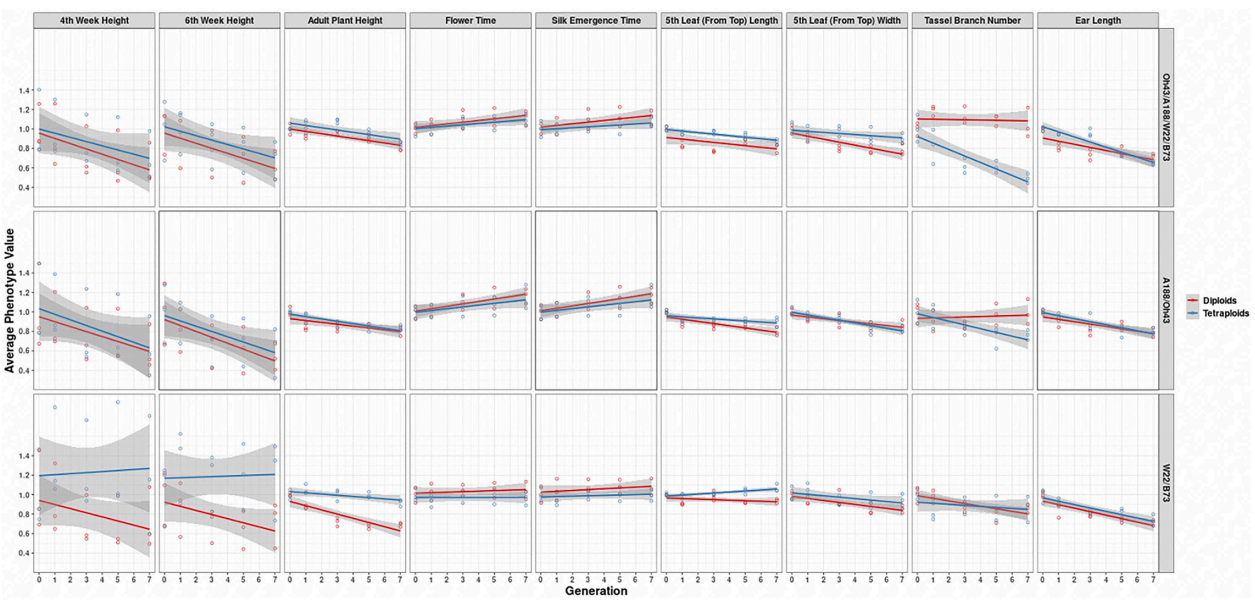


FIGURE 4 | Visualization of the interaction between ploidy and genotype in the 2009 data. Generation (*x*-axis) vs. averaged phenotypic values across biological replicates (*y*-axis) conditioned on the genetic constitution (rows) and trait (columns). Two linear regression lines are superimposed on every panel and their color indicates the diploid (red) and tetraploid plants (blue). The gray-colored band on a regression indicates the 95% confidence band for the whole line.

for this trait. **Figure 5** (2008 data) and **Figure 6** (2009 data) plot the average phenotypic values vs. generation for the traits measured in this case using data for all genotypes and all fields together. Overall, there is not a clear distinction in the diploid and tetraploid comparison.

Because there is not a clear distinction of the inbreeding depression plots between the diploids and tetraploids, we considered the possibility that the chromosomal segregation in the tetraploid might not be as expected. The unlikely scenario that only unlike homologues would pair with each other and that like homologues would proceed to the same pole from such pairing might produce the observed inbreeding curves because that scenario would mimic diploid segregation, although this scenario would still predict a slower inbreeding progression than in diploids, being 1/8 vs. 1/2 at any one locus. To examine this possibility, early, and late meiosis I samples of selected homozygous or heterozygous tetraploid genotypes were examined (**Figure 7**). The patterns of pairing observed indicated the typical array of tetraploid pairing involving bivalents, trivalents, and quadrivalents.

To examine this issue further, chromosomal karyotypes were performed on selfed progeny of tetraploid W22/B73 hybrid

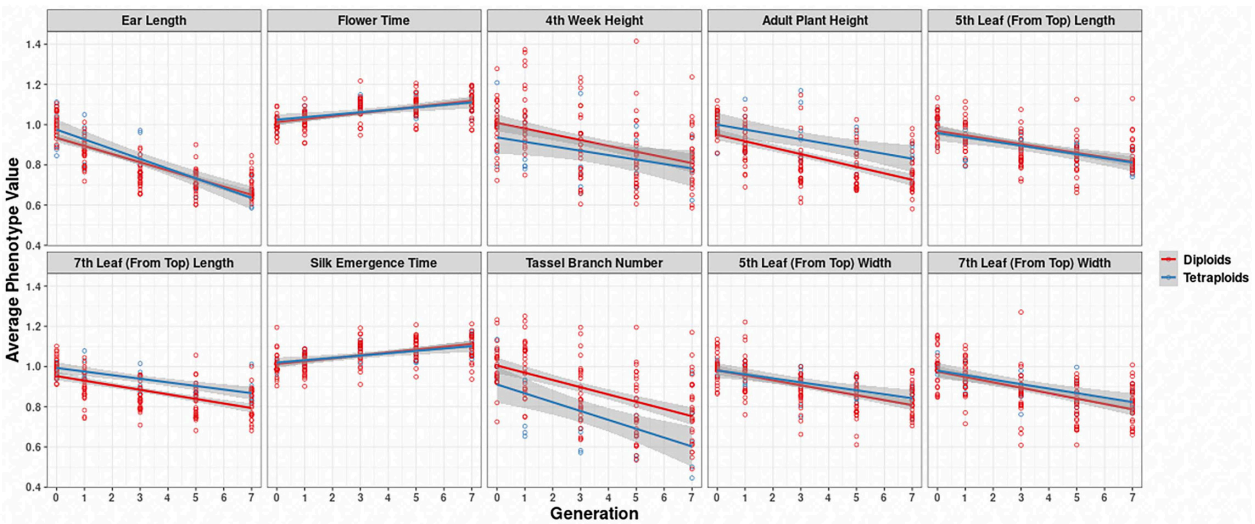


FIGURE 5 | Visualization of inbreeding depression for the 10 phenotypes in the 2008 data conditioned on ploidy. Generation (*x*-axis) vs. averaged phenotypic values across biological replicates (*y*-axis) for the 10 traits (panels). Two linear regression lines are superimposed on every panel, and their color indicates the diploid (red) and tetraploid plants (blue). The gray-colored band on a regression indicates the 95% confidence band for the whole line.

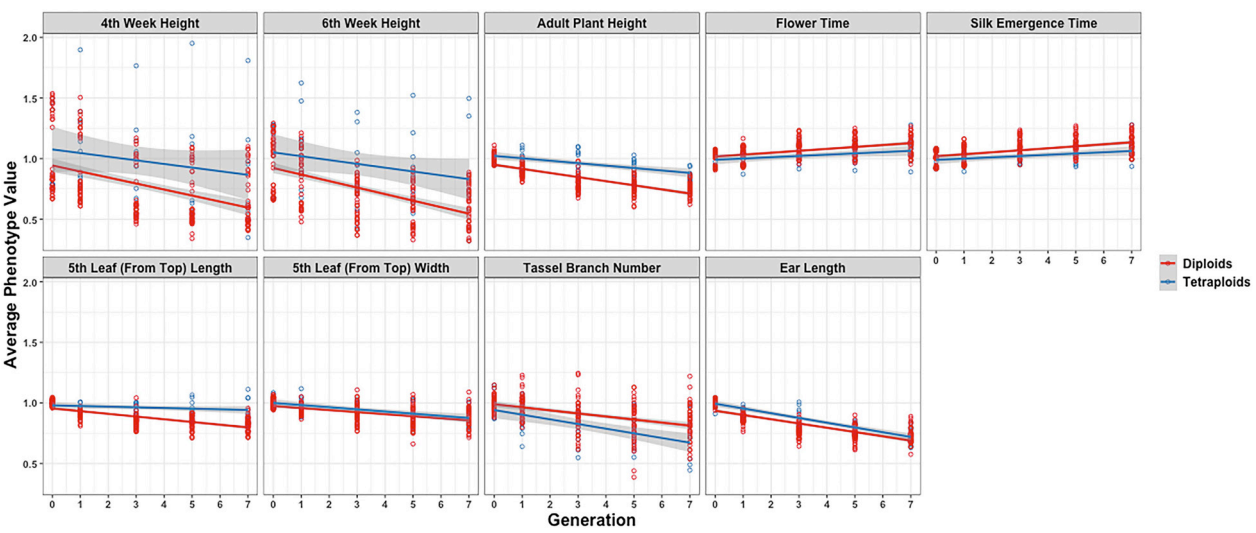


FIGURE 6 | Visualization of inbreeding depression for the nine phenotypes in the 2009 data conditioned on ploidy. Generation (*x*-axis) vs. averaged phenotypic values across biological replicates (*y*-axis) for the 10 traits (panels). Two linear regression lines are superimposed on every panel and their color indicates the diploid (red) and tetraploid plants (blue). The gray-colored band on a regression indicates the 95% confidence band for the whole line.

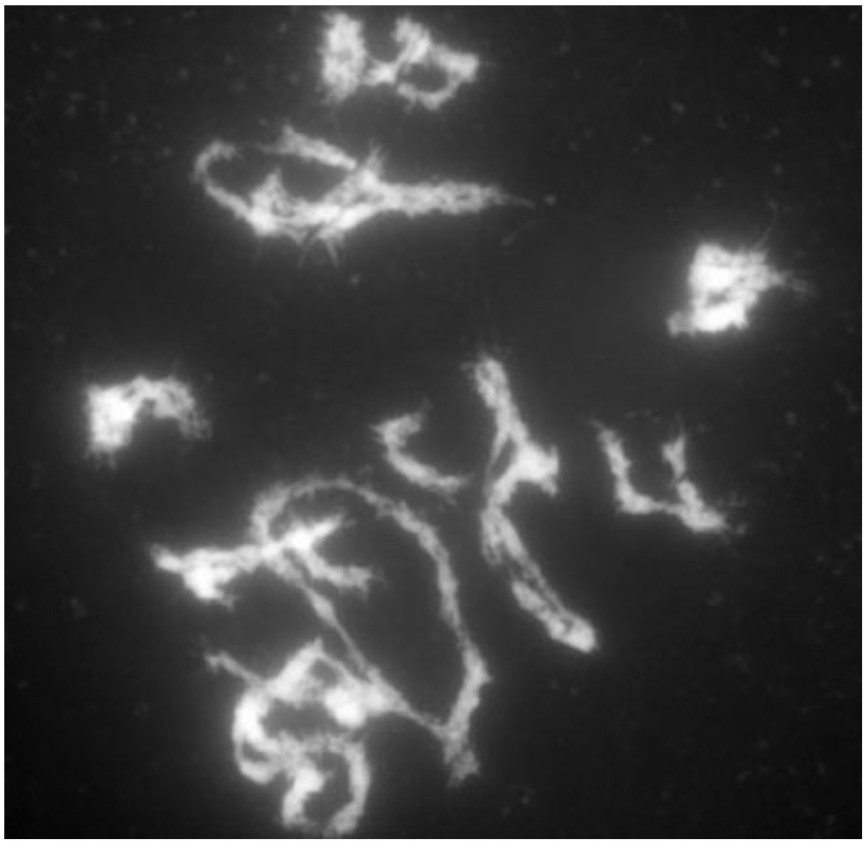


FIGURE 7 | Late prophase of meiosis I in a tetraploid B73. Carmine stain of late prophase of a tetraploid B73 plant. Note the array of bivalents and multivalents.

plants. In this comparison, markers on chromosomes 1, 2, 4, 5, 8, 9, and 10 could be distinguished. Tetraploid maize plants were analyzed using FISH. A cocktail of probes was used to identify the separate chromosomes in the plants as well as the genetic background, W22 or B73, of those chromosomes.

Using the noted chromosomal features, several tetraploid plants were karyotyped. There were 18 plants that were able to be examined. Using the karyotypes, chromosome totals were obtained representing those from the W22 genotype, those from the B73 genotype as well as both genotypes together. A total of 372 chromosomes were assigned to the two genotypes: 193 W22 chromosomes and 179 B73 chromosomes. Chromosome sets that were not complete or could not be assigned to either genotype were excluded from the data set. Chromosome set categories were assigned based on the distribution of W22 chromosomes and B73 chromosomes for each tetraploid chromosome number. The sets were as follows with W22 being the first number and B73 being the second: 4:0, 3:1, 2:2, 1:3, and 0:4. These sets were tallied, and a chi-square test was performed on the observed numbers of each category to test if they followed a distribution of 1:8:18:8:1 predicted from that random joining of gametes in an autotetraploid using the centromere as the marker. There were a total of 92 tallied individual distributions. The observed distributions for the five separate categories were 4:0 = 6, 3:1 = 20, 2:2 = 41, 1:3 = 23, and 0:4 = 2. The chi-square for these data is 5.60 with 4° of freedom and a $p > 0.10$. Thus, the deviation from the predicted segregation for an autotetraploid is not significant.

DISCUSSION

In this study, the pattern of inbreeding depression was compared at the diploid and tetraploid levels in matched genotypes starting at the F1 hybrid stage. The connection among allelic constitutions, genic interactions in the genome, and the phenotype is not understood, including how inbreeding across the genome intersects with these considerations. Nevertheless, there is not a clear distinction that inbreeding depression is slower for tetraploids than diploids. Indeed, for some characteristics, there is an accelerated depression. Various studies of heterosis in diploids have indicated that different characteristics are not necessarily correlated with regard to the magnitude of heterosis (Flint-Garcia et al., 2009; Yao et al., 2013). This relationship is apparent in these data as well. Nevertheless, the overall trend is that there is not an obvious difference between the inbreeding patterns at the two ploidy levels.

On the assumption that inbreeding results from the homozygosis of slightly deleterious recessive alleles, the results are not consistent with this concept. The homozygosis of alleles in an autotetraploid depends on the position of the gene on

the chromosome because recombination between the gene in question and the centromere will allow double reduction to occur, i.e., the entry of the same allele from two homologues to a meiotic end product to produce a homozygous gamete from a heterozygous parent (Levings and Alexander, 1966). However, even for genes near the tips of chromosome arms, homozygosis would not be predicted at the same rate as in a diploid.

The behavior of chromosomes in meiosis I in a representative homozygous and heterozygous genotype was found to exhibit the complicated pattern of pairing typical of autotetraploids (Randolph, 1935) with chromosomes synapsed in pairs but switching pairing partners along the length of the chromosome. The pairing configurations at the end of meiosis I showed pairs, trivalents, and quadrivalents as would be predicted from the pairing associations at the pachytene stage. Distinguishing chromosomal features between genotypes was documented in progeny of a selfed heterozygote and revealed no significant deviation from the expectations of tetraploid frequencies. Thus, there is no reason to suspect from the chromosome behavior that segregation is unusual for the autotetraploids in this study.

The rate of recombination in diploid and tetraploids might potentially impact the rate of inbreeding depression, although there is little known about this issue. The data that are available suggest that the recombination frequency is higher in autotetraploids (Pecinka et al., 2011; Wang and Luo, 2012), but how this would intersect with chromosomal segregation and double reduction to affect homozygosis is not known.

Another factor that could impact the inbreeding curve is that tetraploids will produce many gametes that are heterozygous. There is the potential that heterozygous pollen tubes would grow faster than homozygous ones for certain loci, in other words, exhibit heterosis. Chase (1980) claimed evidence for heterosis in diploid gametophytes. Therefore, if indeed this were the case, the preferential success of heterozygous gametophytes would only be predicted to slow the rate of homozygosis even more because the opportunity to produce homozygous zygotes would be reduced. However, the data from chromosomal feature segregation suggest that there is not a greater number of heterozygous gametes to a measurable degree.

Given that chromosome behavior and heterotic gametophytes favor a slowed progression to homozygosis, what could account for the observed results? Busbice and Wilsie (1966) suggested from work in alfalfa with similar results that a shift in allelic dosage might account for the related inbreeding curves. Indeed, changing allelic dosage is more similar between diploids and tetraploids than homozygosis (Birchler, 2012). As noted earlier, heterosis in triploid hybrids shows evidence of a dosage component and perhaps those results and the ones presented here reflect a related mechanism. Many quantitative traits exhibit semidominance or dosage-sensitive effects (Birchler and Veitia, 2012; Birchler et al., 2016), and this fact raises the possibility that the control of heterosis operates at the level of regulatory interactions (Wang et al., 2015, 2017). Further work is required to understand heterosis in detail, of course, but determining its behavior under as many circumstances as possible will establish the evidence that needs to be explained by a comprehensive framework.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HY conducted the biological experiments. SS performed the statistical analysis. NS performed the segregation analysis. FH performed the meiosis analysis. RD supervised the statistical analysis. JB designed the experiments and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the NSF plant genome grants DBI-0733857 (JB and RD) and IOS-1545780 (JB). SS was supported by the Office of Naval Research (ONR-BAA N000141812741) and the National Science Foundation (DMS-1854667/1854662).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.564928/full#supplementary-material

REFERENCES

Albert, P. S., Gao, Z., Danilova, T. V., and Birchler, J. A. (2010). Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenet. Genome Res.* 129, 6–16. doi: 10.1159/000314342

Alexander, D. E., and Sonnemaker, E. H. (1961). Inbreeding depression in autotetraploid maize. *Maize Genet. Coop. News Lett.* 35:45.

Bingham, E. T. (1980). "Maximizing heterozygosity in autotetraploids," in *Polyploidy: Biological Relevance*, ed. W. H. Lewis (New York, NY: Plenum), 471–489. doi: 10.1007/978-1-4613-3069-1_24

Bingham, E. T., Groose, R. W., Woodfield, D. R., and Kidwell, K. K. (1994). Complementary gene interactions in alfalfa are greater in autotetraploids than diploids. *Crop Sci.* 34, 823–829. doi: 10.2135/cropsci1994.0011183x003400040001x

Birchler, J. A. (2012). "Genetic consequences of polyploidy in plants," in *Polyploidy and Genome Evolution*, eds P. S. Soltis and D. E. Soltis (Heidelberg: Springer).

Birchler, J. A., Johnson, A. F., and Veitia, R. A. (2016). Kinetics genetics: incorporating the concept of genomic balance into an understanding of quantitative traits. *Plant Sci.* 245, 128–134. doi: 10.1016/j.plantsci.2016.02.002

Birchler, J. A., and Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14746–14753. doi: 10.1073/pnas.1207726109

Briggle, L. W. (1963). Heterosis in wheat- a review. *Crop Sci.* 3, 407–412. doi: 10.2135/cropsci1963.0011183x000300050011x

Bruce, A. B. (1910). The Mendelian theory of heredity and the augmentation of vigor. *Science* 32, 627–628. doi: 10.1126/science.32.827.627-a

Busbice, T. H., and Wilsie, C. P. (1966). Inbreeding depression and heterosis in autotetraploids with application to *Medicago sativa* L. *Euphytica* 15, 52–67.

Charlesworth, D., and Willis, J. H. (2009). The genetics of inbreeding depression. *Nat. Rev. Genet.* 10, 783–796.

Chase, S. S. (1980). "Studies of monoploids, diploids and tetraploids of maize in relation to heterosis and inbreeding depression," in *Proceedings of the Argentine Society of Genetics*.

Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14, 471–482. doi: 10.1038/nrg3503

Demarly, Y. (1963). Genetique des tetraploids et amelioration des plants. *Ann. Amelior. Plant* 13, 307–400.

Dewey, D. R. (1966). Inbreeding depression in diploid, tetraploid and hexaploid crested wheatgrass. *Crop Sci.* 6, 144–147. doi: 10.2135/cropsci1966.0011183x000600020011x

Dewey, D. R. (1969). Inbreeding depression in diploid and induced-autotetraploid crested wheatgrass. *Crop Sci.* 9, 592–595. doi: 10.2135/cropsci1969.0011183x000900050023x

Doyle, G. G. (1979). The allotetraploidization of maize: part I: the physical basis—differential pairing affinity. *Theor. Appl. Genet.* 54, 103–112. doi: 10.1007/bf01159463

Doyle, G. G. (1986). The allotetraploidization of maize: 4. Cytological and genetic evidence indicative of substantial progress. *Theor. Appl. Genet.* 71, 585–594. doi: 10.1007/bf00264261

Dunbier, M. W., and Bingham, E. T. (1975). Maximum heterozygosity in alfalfa: results using haploid-derived autotetraploids. *Crop Sci.* 15, 527–531.

Duvick, D. N. (1999). "Heterosis: feeding people and protecting natural resources," in *Genetics and Exploitation of Heterosis in Crops*, eds J. G. Coors and S. Pandey (Madison, WI: American Society of Agronomy, Inc. and Crops Science Society of America, Inc.), 19–29. doi: 10.2134/1999.geneticsandexploitation.c3

East, E. M. (1936). Heterosis. *Genetics* 21, 375–397.

Flint-Garcia, S. A., Buckler, E. S., Tiffin, P., Ersoz, E., and Springer, N. M. (2009). Heterosis is prevalent for multiple traits in diverse maize germplasm. *PLoS One* 4:e7433. doi: 10.1371/journal.pone.0007433

Gallais, A. (1984). An analysis of heterosis vs. inbreeding effects with an autotetraploid cross-fertilized plant: *Medicago sativa* L. *Genetics* 106, 123–137.

Groose, R. W., Talbert, L. E., Kojis, W. P., and Bingham, E. T. (1989). Progressive heterosis in autotetraploid alfalfa: studies using two types of inbreds. *Crop Sci.* 29, 1173–1177. doi: 10.2135/cropsci1989.0011183x002900050015x

Jones, D. F. (1917). Dominance of linked factors as a means of accounting for heterosis. *Genetics* 2, 466–479.

Kato, A., and Birchler, J. A. (2006). Induction of tetraploid derivatives of maize inbred lines by nitrous oxide gas treatment. *J. Hered.* 97, 39–44. doi: 10.1093/jhered/esj007

Kato, A., Lamb, J. C., and Birchler, J. A. (2004). Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13554–13559. doi: 10.1073/pnas.0403659101

Levings, C. S. III, and Alexander, D. E. (1966). Double reduction in autotetraploid maize. *Genetics* 54, 1297–1305.

Levings, C. S., Dudley, J. W., and Alexander, D. E. (1967). Inbreeding and crossing in autotetraploid maize. *Crop Sci.* 7, 72–73. doi: 10.2135/cropsci1967.0011183x000700010025x

Li, X., and Brummer, C. (2009). Inbreeding depression for fertility and biomass in advanced generations of inter-and intrasubspecific hybrids of tetraploid alfalfa. *Crop Sci.* 49, 13–19. doi: 10.2135/cropsci2008.04.0205

Li, Z., Li, B., and Tong, Y. (2008). The contribution of distant hybridization with decaploid *Agropyron elongatum* to wheat improvement in China. *J. Genet. Genomics* 35, 451–456. doi: 10.1016/s1673-8527(08)60062-4

Mok, D. W. S., and Peloquin, S. J. (1975). Breeding value of 2n pollen (diploandroids) in tetraploid x diploid crosses in potato. *Theor. Appl. Genet.* 46, 307–314. doi: 10.1007/bf00281153

Pecinka, A., Fang, W., Rehsmeier, M., Levy, A. A., and Mittelsten-Scheid, O. (2011). Polyploidization increases meiotic recombination frequency in *Arabidopsis*. *BMC Biol.* 9:24. doi: 10.1186/1741-7007-9-24

Randolph, L. F. (1935). Cytogenetics of tetraploid maize. *J. Agric. Res.* 50, 591–606.

Rice, J. S., and Dudley, J. W. (1974). Gene effects responsible for inbreeding depression in autotetraploid maize. *Crop Sci.* 14, 390–393. doi: 10.2135/cropsci1974.0011183x001400030015x

Riddle, N. C., and Birchler, J. A. (2008). Comparative analysis of inbred and hybrid maize at the diploid and tetraploid levels. *Theor. Appl. Genet.* 116, 563–576. doi: 10.1007/s00122-007-0691-1

Riddle, N. C., Jiang, H., An, L., Doerge, R. W., and Birchler, J. A. (2010). Gene expression analysis at the intersection of ploidy and hybridity in maize. *Theor. Appl. Genet.* 120, 341–353. doi: 10.1007/s00122-009-1113-3

Riddle, N. C., Kato, A., and Birchler, J. A. (2006). Genetic variation for the response to ploidy change in *Zea mays* L. *Theor. Appl. Genet.* 114, 101–111. doi: 10.1007/s00122-006-0414-z

Shull, G. H. (1908). The composition of a field of maize. *Am. Breeders Assoc. Rep.* 4, 296–301. doi: 10.1093/jhered/os-4.1.296

Sockness, B. A., and Dudley, J. W. (1989a). Performance of single and double cross autotetraploid maize hybrids with different levels of inbreeding. *Crop Sci.* 29, 875–879. doi: 10.2135/cropsci1989.0011183x002900040006x

Sockness, B. A., and Dudley, J. W. (1989b). Morphology and yield of isogenic diploid and tetraploid maize inbreds and hybrids. *Crop Sci.* 29, 1029–1032. doi: 10.2135/cropsci1989.0011183x002900040041x

Tan, C., Pan, Q., Cui, C., Xiang, Y., Ge, X., and Li, Z. (2016). Genome-wide gene/genome dosage imbalance regulates gene expressions in synthetic *Brassica napus* and derivatives (AC, AAC, CCA, CCAA). *Front. Plant Sci.* 7:1432. doi: 10.3389/fpls.2016.01432

Tysdal, H. M., Kiesselbach, T. A., and Westover, H. L. (1942). Alfalfa breeding. *Nebraska Agric. Exp. Stn. Res. Bull.* 124, 1–46.

Wang, L., Greaves, I. K., Groszmann, M., Wu, L. M., Dennis, E. S., and Peacock, W. J. (2015). Hybrid mimics and hybrid vigor in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4959–E4967.

Wang, L., and Luo, Z. (2012). Polyploidization increases meiotic recombination frequency in *Arabidopsis*: a close look at statistical modeling and data analysis. *BMC Biol.* 10:30. doi: 10.1186/1741-7007-10-30

Wang, L., Wu, L. M., Greaves, I. K., Zhu, A., Dennis, E. S., and Peacock, W. J. (2017). PIF4-controlled auxin pathway contributes to hybrid vigor in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 114, E3555–E3562.

Washburn, J. D., McElfresh, M. J., and Birchler, J. A. (2019). Progressive heterosis in genetically defined tetraploid maize. *J. Genet Genomics* 46, 389–396.

Welch, J. E. (1962). Linkage in autotetraploid maize. *Genetics* 47, 367–396.

Williams, R. D. (1931). Self- and cross-fertility and flowering habits of certain herbage grasses and legumes. *Welsh Plant Breeding Stn. Bull.* 12, 217–220.

Yao, H., Dogra-Gray, A., Auger, D. L., and Birchler, J. A. (2013). Genomic dosage effects on heterosis in triploid maize. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2665–2669.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fgene.2021.589160

Gene and Transposable Element Expression Evolution Following Recent and Past Polyploidy Events in *Spartina* (Poaceae)

*Delphine Giraud[1], Oscar Lima[1], Mathieu Rousseau-Gueutin[2], Armel Salmon[1] and Malika Aïnouche[1]**

[1] UMR CNRS 6553 Ecosystèmes, Biodiversité, Evolution (ECOBIO), Université de Rennes 1, Rennes, France, [2] IGEPP, INRAE, Institut Agro, Univ Rennes, Le Rheu, France

Gene expression dynamics is a key component of polyploid evolution, varying in nature, intensity, and temporal scales, most particularly in allopolyploids, where two or more sub-genomes from differentiated parental species and different repeat contents are merged. Here, we investigated transcriptome evolution at different evolutionary time scales among tetraploid, hexaploid, and neododecaploid *Spartina* species (Poaceae, Chloridoideae) that successively diverged in the last 6–10 my, at the origin of differential phenotypic and ecological traits. Of particular interest are the recent (19th century) hybridizations between the two hexaploids *Spartina alterniflora* (2*n* = 6*x* = 62) and *S. maritima* (2*n* = 6*x* = 60) that resulted in two sterile F1 hybrids: *Spartina* × *townsendii* (2*n* = 6*x* = 62) in England and *Spartina* × *neyrautii* (2*n* = 6*x* = 62) in France. Whole genome duplication of *S.* × *townsendii* gave rise to the invasive neo-allododecaploid species *Spartina anglica* (2*n* = 12*x* = 124). New transcriptome assemblies and annotations for tetraploids and the enrichment of previously published reference transcriptomes for hexaploids and the allododecaploid allowed identifying 42,423 clusters of orthologs and distinguishing 21 transcribed transposable element (TE) lineages across the seven investigated *Spartina* species. In 4*x* and 6*x* mesopolyploids, gene and TE expression changes were consistent with phylogenetic relationships and divergence, revealing weak expression differences in the tetraploid sister species *Spartina bakeri* and *Spartina versicolor* (<2 my divergence time) compared to marked transcriptome divergence between the hexaploids *S. alterniflora* and *S. maritima* that diverged 2–4 mya. Differentially expressed genes were involved in glycolysis, post-transcriptional protein modifications, epidermis development, biosynthesis of carotenoids. Most detected TE lineages (except *SINE* elements) were found more expressed in hexaploids than in tetraploids, in line with their abundance in the corresponding genomes. Comparatively, an astonishing (52%) expression repatterning and deviation from parental additivity were observed following recent reticulate evolution (involving the F1 hybrids and the neo-allododecaploid *S. anglica*), with various patterns of biased homoeologous gene expression, including genes involved in epigenetic regulation. Downregulation of TEs was observed in both

OPEN ACCESS

Edited by:
Yves Van de Peer, Ghent University, Belgium

Reviewed by:
Pamela Soltis, University of Florida, United States
Clayton J. Visger, California State University, Sacramento, United States

***Correspondence:**
Malika Aïnouche
malika.ainouche@univ-rennes1.fr

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *30 July 2020*
Accepted: *23 February 2021*
Published: *25 March 2021*

Citation:
Giraud D, Lima O, Rousseau-Gueutin M, Salmon A and Aïnouche M (2021) Gene and Transposable Element Expression Evolution Following Recent and Past Polyploidy Events in Spartina (Poaceae). Front. Genet. 12:589160. doi: 10.3389/fgene.2021.589160

hybrids and accentuated in the neo-allopolyploid. Our results reinforce the view that allopolyploidy represents springboards to new regulatory patterns, offering to worldwide invasive species, such as *S. anglica,* the opportunity to colonize stressful and fluctuating environments on saltmarshes.

Keywords: allopolyploidy, hybridization, transcriptome evolution, transposable elements (TE), *Spartina*

INTRODUCTION

Gene expression dynamics is a key component of polyploid evolution, and has received considerable interest in the last decades, where studies on various polyploid systems revealed an important role of whole genome duplication (WGD) in modulating diverse and novel gene expression patterns (Chen, 2007; Flagel et al., 2008; Yoo et al., 2014). Comparative analyses indicate that the evolution of duplicated gene expression has a temporal dimension, and that immediate responses to polyploidy may not only last over long periods, but also contribute to the long-term processes of diploidization and fractionation (Wendel et al., 2018). One of the key parameters of this dynamics is the divergence level between the genomes being merged and duplicated during the polyploid speciation process (Buggs et al., 2012; Tayalé and Parisod, 2013). Neopolyploids are usually classified as autopolyploids, where homologous genomes (i.e., within species) are duplicated, or as allopolyploids involving duplication of more or less divergent (homoelogous) genomes reunited in the same nucleus following interspecific hybridization (Stebbins, 1971; reviewed in Doyle et al., 2008). Most autopolyploids so far explored exhibit moderate transcriptome or proteome alteration compared to their diploid progenitors (Albertin et al., 2005; Parisod et al., 2010b; del Pozo and Ramirez-Parra, 2014; Visger et al., 2019; Song et al., 2020). In contrast, allopolyploidy seems to be accompanied by profound parental expression repatterning in naturally formed neopolyploids (Hegarty et al., 2006; Chelaifa et al., 2010b; Buggs et al., 2011), experimentally resynthesized allopolyploids and/or their naturally established counterparts (Wendel, 2000; Chen and Ni, 2006; Ha et al., 2009; Soltis et al., 2015), such as in oilseed rape, wheat, cotton, or coffea (Akhunova et al., 2010; Higgins et al., 2012; Combes et al., 2013; Yoo et al., 2013; Wu et al., 2018). These dynamics reflect both the effects of the reunion of divergent regulatory networks (resulting from hybridization) and the effects of genetic redundancy (resulting from WGD), all these contributing to a "transcriptomic shock" (Hegarty et al., 2006), as a functional extension of what was earlier termed as "genomic shock" (McClintock, 1984), including its associated genetic and epigenetic regulatory processes (Hu and Wendel, 2019).

Duplicated genes may undergo partitioning of ancestral functions (subfunctionalization, Force et al., 1999) or benefit from mutations conferring new functionality (neofunctionalization, Ohno, 1970), which in turn affects the long-term retention of the duplicated copies (duplication-degenerescence-complementation model, Lynch and Force, 2000). The enhanced functional plasticity resulting from gene copy redundancy and allelic diversity facilitates the emergence of novel variation, traits and phenotypes (Finigan et al., 2012), thus contributing to species adaptation and long-term diversification (Comai, 2005; Conant and Wolfe, 2008; Jackson and Chen, 2010; Madlung, 2013; Tank et al., 2015; Van de Peer et al., 2017).

Expression evolution in polyploids has been extensively explored by testing their deviation from an expected transcriptomic "parental additivity." This non-additive parental expression may be perceived by considering either the overall gene expression level (usually measured by comparisons with the average expression of both parental species, i.e., Mid-Parent Value; MPV) or by considering the relative contribution of each homoeologous copy to the total expression level (Grover et al., 2012). The high frequency of biases in the respective contributions of homeologs reported in various allopolyploids lead to the "genomic dominance" concept whereby one parental (homoeologous) subgenome expression state is exhibited in strong preference over the other parental expression state (Flagel and Wendel, 2009).

Recent studies provide accumulating evidence that the "dominant" subgenome is early established following hybridization and may be maintained over generations (Edger et al., 2017; Bird et al., 2018). This, very interestingly, permits to connect the short-term genome dominance phenomenon to the long-term fractionation process affecting polyploid genomes (Wendel et al., 2018), where biases in gene loss between duplicated homoeologous genomes result from the selection against loss of the most expressed copy (Schnable et al., 2011). The corollary is that following allopolyploidy, the "dominant" subgenome (the highest expressed) is more likely to be retained than the "recessive" subgenome (the lowest expressed) that becomes the most fractionated (Cheng et al., 2018). As expected, subgenome dominance is not observed in neo-autopolyploids (Garsmeur et al., 2014) suggesting a major role of the composition of the two subgenomes being merged in this process, notably in their content of transposable or regulatory elements.

Allopolyploidy merges and duplicates more or less differentiated genomes, including repetitive sequences that represent a dynamic component of plant genomes (Bennetzen and Wang, 2014). Transposable elements (TEs) affect in various ways the structure and expression of polyploid genomes. Several studies (reviewed in Vicient and Casacuberta, 2017) have reported transcriptional reactivation of some TE lineages in hybrids and allopolyploids or new insertions as found in tobacco, sunflower, wheat, or *Brachiaria* sp. (Ungerer et al., 2006; Petit et al., 2010; Yaakov and Kashkush, 2012; Santos et al., 2015). As TE expression is controlled by epigenetic regulation (e.g., DNA hypermethylation and siRNAs), new TE regulation

following polyploid speciation may result in altered neighbor gene expression including gene silencing and novel gene expression patterns (Kashkush et al., 2003; Hollister and Gaut, 2009; Parisod et al., 2010a; Zhang et al., 2015; Lerat et al., 2019). In allopolyploids, divergence between the parental genomes in terms of TE abundance and distribution is expected to increase subgenome dominance, as the TE-rich subgenome is expected to be targeted by epigenetic regulation and then less expressed than the TE-sparse subgenome (Bird et al., 2018; Bottani et al., 2018).

This complex interplay between parental genome legacy, the challenges faced by the needed coordination of divergent regulatory networks in the same nucleus and the connected processes taking place at the early and latest stages of polyploid species evolution raise critical questions about the rules and mechanisms shaping modern extant genomes and their predictability among polyploid lineages.

The *Spartina* genus (Poaceae, Chloridoideae) provides excellent opportunities to explore such processes at different evolutionary time scales, providing a phylogenetically well-understood framework of polyploidization that occurred in the last 10–12 my (Ainouche et al., 2012; Rousseau-Gueutin et al., 2015). Hybridization and polyploidy have recurrently shaped the history of this genus (also considered as a subsection in the *Sporobolus* genus; Peterson et al., 2014 but see Bortolus et al., 2019) and have resulted in modern species with various ploidy, ranging from tetraploids to dodecaploids, with a basic chromosome number of $x = 10$ (Marchant, 1968a; Ainouche et al., 2012). No diploid species was reported to date in this genus which most likely evolved from a tetraploid ancestral lineage ($2n = 4x = 40$) that diversified mainly in North and South America (and more recently in Europe) where they colonize coastal and inland marshes (Mobberley, 1956; Baumel et al., 2002a). In this study, we focused on seven species representing different ploidy, selected on the basis of their known phylogenetic history (Baumel et al., 2002a; Fortune et al., 2007; Rousseau-Gueutin et al., 2015). In tetraploids, two sister species were selected: *S. bakeri* and *Spartina patens*. *S. bakeri* (sand cordgrass) is native to the southeastern United States, where it grows on sandy beaches along the coast and in inland freshwater habitat in Florida and southern Georgia. It is morphologically highly similar to *S. patens* (saltmeadow grass), from which it can be distinguished by its unique vegetative habit (wanting rhizomes) and tolerance to freshwater (Mobberley, 1956). *S. patens* is distributed along the eastern United States to the Caribbean and northeast Mexico and has been introduced to the Mediterranean coast since the mid-19th century (Fabre, 1849) where it has been first considered as an endemic species and named *S. versicolor* Fabre. Recent molecular studies revealed that this latter actually represents an introduced, invasive genotype of *S. patens* (Prieto et al., 2011; Baumel et al., 2016). In the hexaploid clade, which diverged from the tetraploids 6–10 mya (Rousseau-Gueutin et al., 2015), we analyzed two species that diverged c.a. 2–4 mya: *S. alterniflora* ($2n = 6x = 62$) native to the Atlantic American coasts and *S. maritima* ($2n = 6x = 60$) native to the European Atlantic coast (Baumel et al., 2002a). During the 19th century, introductions of *S. alterniflora* to Europe led to hybridization with the native species and had important ecological and evolutionary consequences (reviewed in Ainouche et al., 2009; Strong and Ayres, 2013). Hybridization with *S. maritima* (as male parent) resulted in two natural F1 hybrids: *S.* × *townsendii* ($2n = 6x = 62$; Groves and Groves, 1880) in Southern England and *S.* × *neyrautii* ($2n = 6x = 62$; Foucaud, 1897) in south–west France (Marchant, 1963; Baumel et al., 2003; Ainouche and Wendel, 2014). Chromosome doubling of *S.* × *townsendii* at the end of the 19th century resulted in the formation of the new fertile allododecaploid ($2n = 12x = 124$) *S. anglica* (Hubbard, 1965; Marchant, 1968a). This system is now considered as one of the textbook examples of recent allopolyploid speciation, which allows examining the immediate effects of hybridization and genome duplication in natural populations (Ainouche et al., 2004).

Polyploidy and hybridization have important adaptive consequences in *Spartina*. Hexaploid species evolved anatomical and physiological traits [including salt and xenobiotic tolerance increased by specific cells such as salt glands or trichomes, production of DMSP (Dimethylsulfoniopropionate), an osmoprotectant molecule; Céccoli et al., 2015; Paredes-Páliz et al., 2017; Robertson et al., 2017; Rousseau et al., 2017; Cavé-Radet et al., 2019b] providing the ability to colonize low marsh zones and tolerate several hours of tidal immersion (Maricle et al., 2009; Bedre et al., 2016). Invasive abilities appear as an immediate consequence of hybridization and/or genome duplication in *Spartina* (Strong and Ayres, 2013; Ainouche and Gray, 2016). The neopolyploid *S. anglica* has rapidly expanded in range in its native region (Europe) and has been deliberately introduced in several continents for land reclamation and estuary stabilization (e.g., China, Australia, and New Zealand), where several attempts of eradication are being conducted to control the species (Gray and Benham, 1990; Hammond and Cooper, 2002; Ainouche et al., 2009). This recently formed allopolyploid species exhibits higher ecological amplitude and higher stress tolerance than its parental species (Cavé-Radet et al., 2019b). The highly duplicated nature of its allododecaploid genome is thought to play an important role in the plasticity and the observed adaptive features. First investigations of global gene expression using microarrays on this system revealed various patterns of non-additive parental expression affecting genes involved in stress tolerance (Chelaifa et al., 2010b; Ferreira de Carvalho et al., 2017). Significant DNA methylation changes were also observed notably in regions flanking transposable elements, as evaluated from methylation-sensitive AFLP (Salmon et al., 2005) and methylation sensitive transposon display (Parisod et al., 2009). More recently, reference transcriptomes were built using Next Generation Sequencing for the hexaploid and allododecaploid species (Ferreira de Carvalho et al., 2013b; Boutte et al., 2016) and small-RNAs (Cavé-Radet et al., 2019a) and repetitive sequences (Giraud et al., 2020) from tetraploid and hexaploid *Spartina* genomes were annotated.

In this study, gene and TE expressions were analyzed at different evolutionary time scales (**Figure 1**). The two

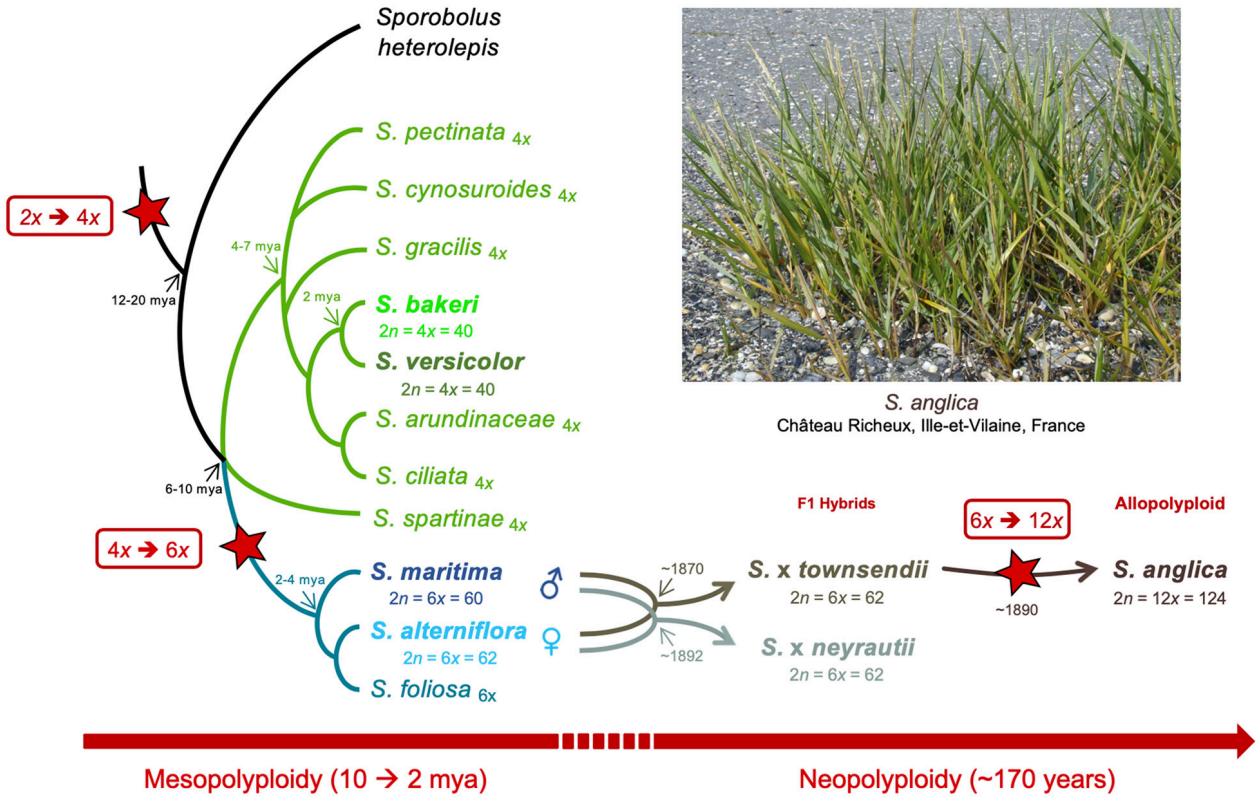


FIGURE 1 | Phylogenetic relationships among *Spartina* species (redrawn from Baumel et al., 2002a), and position of the analyzed tetraploid (*Spartina bakeri* and *Spartina versicolor*), hexaploid (*Spartina maritima*, *Spartina alterniflora*, *Spartina* × *townsendii*, *Spartina* × *neyrautii*) and dodecaploid (*Spartina anglica*) taxa analyzed in this study. Divergence times (my) represent upper limits estimated so far (Rousseau-Gueutin et al., 2015). Stars indicate whole genome duplication (WGD) events during *Spartina* evolution.

sister tetraploid species and two hexaploid species were first compared to understand how gene and TE expressions evolved since their divergence 6–10 mya. Then, the evolution of gene and TE expression following recent hybridization and neopolyploidy was analyzed by comparing the species involved in the recent allopolyploid speciation, including the hexaploid parental species, their reciprocal F1 hybrids and the allododecaploid formed during the end of the 19th century. More specifically, we addressed the following questions: (i) how similar are the tetraploid and hexaploid transcriptomes? (ii) how additive are F1 hybrids/allopolyploid transcriptomes compared to parental transcriptomes? (iii) how has the hybridization *vs.* genome duplication affected genes and the expression of TEs? (iv) are the transcriptomes of two independently formed hybrids affected in the same way? (v) what are the functions affected by interspecific hybridizations and/or genome duplication? To answer these questions, reference transcriptomes were assembled for each investigated species (providing new reference transcriptomes for the tetraploid species and enriching previously assembled transcriptomes of hexaploids and hybrid/allopolyploid species Ferreira de Carvalho et al., 2013b; Boutte et al., 2016). Homologous (putative orthologous) transcripts were identified among species and the relative contribution of homeologs was assessed for a subset of non-additively expressed genes in the hybrids and/or the allopolyploid.

MATERIALS AND METHODS

Plant Material, RNA Extraction, and Sequencing

Among the seven *Spartina* species analyzed in this study (**Figure 1**), two individuals per species were collected from their natural habitat along the French coasts for *S. maritima*, *S. alterniflora*, *S.* × *neyrautii*, *S. anglica,* and *S. versicolor*, in Hythe (England) for *S.* × *townsendii* and in Florida (United States) for *S. bakeri* (**Table 1**). Previous transcriptome analyses (Ferreira de Carvalho et al., 2017) indicated that low intraspecific expression variation was encountered among plants collected along a tidal gradient in *Spartina* populations, but in order to limit potential plastic responses among samples, all individuals were transplanted to the same natural site (Morbihan, France on June, 19, 2013) in order to maintain all plants in the same environmental conditions. After 1 month of acclimation (2013/07/18), newly formed leaves from each individual were harvested between 10–12 am and flash frozen in liquid nitrogen. Thereafter, their RNAs

TABLE 1 | Studied polyploid *Spartina* species and summary of RNA-Seq Illumina datasets used.

Species	Synonyms	Ploidy, number of chromosomes	Sample origin	Number of replicates	Total library depth (cumulative number of paired reads)
S. maritima Curtis	***Sporobolus maritimus*** (Curtis) P.M.Peterson & Saarela, comb. nov.	6*x* = 60	Le Hezo (Morbihan, France)	4	146,438,783
S. alterniflora Loisel.	***Sporobolus alterniflorus*** (Loisel.) P.M.Peterson & Saarela, comb. nov.	6*x* = 62	Le Faou (Finistère, France)	4	179,792,458
S. x townsendii H. Groves & J. Groves	***Sporobolus* × *townsendii*** (H. Groves & J. Groves) P.M.Peterson & Saarela, comb. nov.	6*x* = 62	Hythe (Hampshire, United Kingdom)	4	171,773,818
S. × *neyrautii* Foucaud	–	6*x* = 62	Hendaye (Pyrénées Atlantiques, France)	3	139,586,516
S. anglica C. E. Hubb.	***Sporobolus anglicus*** (C.E.Hubb.) P.M.Peterson & Saarela, comb. nov.	12*x* = 124	Le Hezo (Morbihan, France)	8	296,374,022
S. versicolor Fabre	***Sporobolus versicolor*** (Fabre) P.M.Peterson & Saarela, comb. nov.	4*x* = 40	Aresquiers (Hérault, France)	3	76,965,982
S. bakeri Merr.	***Sporobolus bakeri*** (Merr.) P.M.Peterson & Saarela, comb. nov.	4*x* = 40	Florida (United States)	2	85,839,501

were directly extracted according to Chelaifa et al. (2010a). Illumina libraries were prepared according to the Tru-Seq PCR-Free Protocol provided by Illumina. One Illumina library was prepared per individual (two individuals per species) and was sequenced on two independent lanes. The *S. anglica* libraries were sequenced twice more than the hexaploid species because of its larger genome (each individual was sequenced on 4 independent lanes; see **Table 1** and **Supplementary Table 1**). Sequencing was performed at the BGI (Beijing Genomics Institute) using the Hi-Seq 2000 technology (100 bp paired-ends reads; insert-size of 500 bp). Raw sequence data have been deposited at GenBank under the accession number PRJNA657699.

Enrichment of Hexaploid Reference Transcriptomes and New Tetraploid Transcriptomes

Prior to differential gene expression analyses, the reference transcriptomes previously obtained by co-assembling pyrosequencing (Ferreira de Carvalho et al., 2013b) and Illumina data (Boutte et al., 2016) were enriched with the set of newly sequenced reads to build updated references for hexaploid and allododecaploid species (data used for reference transcriptome assemblies are available on GenBank, BioProject PRJNA338100). Considering the complex assembly process of transcriptomes in polyploid species, bioinformatics tools used were configured in order to obtain contigs representing consensus sequences of all expressed homeologous copies (or homeologs) from one given gene (Ferreira de Carvalho et al., 2013b). The expression level estimates for each contig was thus the sum of expressions of homeologs (and possibly, slightly divergent paralogs, resulting from recent individual gene duplication). Individual copies of a given gene were detected after remapping reads to these contigs, detecting SNPs (Single Nucleotide Polymorphisms) and phasing them to build haplotypes (see section Parental Origin of Transgressively Expressed Orthologs).

First, mappings of new Illumina reads on reference transcriptomes were performed using Bowtie2 (parameters: local alignment; authorized mismatch; G, 52, 8; Langmead and Salzberg, 2012). Approximately, 60% of reads per library were aligned on respective transcriptoms. Non-mapped reads, corresponding to the expression of non-assembled transcripts, were then pooled by species and assembled following a method previously described (Boutte et al., 2016). Briefly, reads were assembled using the Trinity algorithm (v. 2.8.4; Grabherr et al., 2011) with a *k*-mer size of 25 bp and a minimum assembled contig length of 48 bp. To improve the assembly, all new contigs with a length higher than 100 bp were co-assembled with Newbler software (parameters: ml = 40 bp; mi = 90%; v.2.8 Margulies et al., 2005). Finally, a selfBLAST of all new contigs was performed using BLAST+ (v. 2.5.0; Altschul et al., 1990) to identify redundant sequences. Contigs included in other contigs were removed and contigs with a minimum overlapping of 50 bp, with an identity higher or equal to 90% were assembled using custom python scripts (Boutte et al., 2016). The newly assembled contigs were then added to previous reference transcriptomes (Boutte et al., 2016) for each hexaploid species (*S. maritima*, *S. alterniflora*, *S.* × *townsendii,* and *S.* × *neyrautii*) and the allopolyploid *S. anglica*. As no reference transcriptomes were available for the analyzed tetraploid species, Illumina libraries sequenced for expression analyses were used to assemble *de novo* reference transcriptomes for each tetraploid species. Reads were assembled using the method described above (Trinity assembly followed by Newbler co-assembly and custom scripts to remove redundancy). Quality of assemblies was tested using BUSCO (v. 4.1.4; Seppey et al., 2019). Transcriptome completeness was measured by comparing new assembled transcriptomes with pre-selected BUSCO genes specific to Poales species (lineage dataset: poales_odb10; creation date: 2020-08-05; number of species: 12; number of BUSCO genes: 4,896; *e*-value < 1e-06).

Functional Annotations

Prior to contig annotations, potential contamination during RNA extraction and/or library sequencing was verified using BLASTn (Altschul et al., 1990) against the NCBI nucleotide database (accessed 08/12/2018). Such contigs displaying similarities with Prokaryote, Fungi, Protist or Metazoan sequences were removed from the analyses (*e*-value < 1e-6 and 90% of identity).

New assembled and reference transcriptomes were annotated as detailed in Ferreira de Carvalho et al. (2013b) and Boutte et al. (2016). Contigs were aligned via BLASTx analyses (one-way Best Blast Hits with an *e*-value < 1e-6) against protein databases from six species: *Arabidopsis thaliana* (genome version TAIR10), *Oryza sativa* (v. IRGSP-1.0), *Brachypodium distachyon* (v. 3.0), *Panicum hallii* (v. 3.1), *Sorghum bicolor* (v. 3), *Zea mays* (v. B73 RefGen v4). Gene Ontologies (GO terms) of each gene identified in our transcriptomes were retrieved from plant protein databases and included in our annotations. Similarities with known protein domains were also found using hmmer [a Hidden Markov Model (HMM) algorithm hmmer.org, v. 3.2] against the Pfam database (v. 32.0 Finn et al., 2014). Figures representing results from annotation analyses were drawn using the VennDiagram (v. 1.6.20; Chen and Boutros, 2011) and the UpSetR v. 1.4 R packages (Conway et al., 2017).

To complete annotations, contigs were also aligned against the *Spartina* transposable element databases built by Giraud et al. (2020) as well as against ribosomal DNA and chloroplast genomes previously assembled (Boutte et al., 2015; Rousseau-Gueutin et al., 2015) using BLASTn (*e*-value < 1e-6; Altschul et al., 1990) and Bowtie2 (parameters: local alignment; authorized mismatch; G, 52, 2; Langmead and Salzberg, 2012).

Identification of Orthologous Transcripts/Regions Among Species

To compare gene expression among *Spartina* species, the OrthoVenn2 program (Xu et al., 2019) was used to identify orthologous contigs or (orthologs) in the different transcriptomes. This program provides clusters of orthologous sequences by sequence comparison analyses (BLASTp) and graph-based clustering among up to eight plant species simultaneously. In order to reduce the computational time, only annotated contigs were conserved for analyses. Prior to OrthoVenn2 analysis, contigs were translated into protein sequences (in the 6 reading frames) using the EMBOSS transeq tool (v. 6.6.0; Madeira et al., 2019). OrthoVenn2 was run with default parameters (*e*-value 1e-5; inflation value 1.5) to identify orthologs among the different transcriptomes. OrthoVenn2 allowed grouping orthologs found in the studied species into the same cluster. Within each cluster, regions or windows where orthologs showed sequence similarities were delimited in order to measure expression levels in each species from the same sequence length. To achieve this, all contigs or orthologs within the same cluster were aligned against each other using BLASTn (*e*-value < 1e-6; Altschul et al., 1990) and windows displaying a maximal number of aligned contigs and a maximal size were selected using custom python scripts for differential expression analyses.

Differential Expression Analyses

Gene expression levels were measured by mapping reads of each sequenced library on newly assembled reference transcriptomes using the Bowtie2 software with the following parameters: local alignment; authorized mismatch; G, 52, 8 corresponding to 90% of minimum identity (Langmead and Salzberg, 2012). For each cluster of orthologs, gene expression levels were given by the number of reads mapped on contigs and included in windows or regions previously selected.

Expression of TEs was also estimated according to the same method as used for gene expression. RNA-seq libraries were mapped using Bowtie2 (same settings as mentioned above) on their respective *Spartina* TE database previously built by Giraud et al. (2020). The expression of each TE family was measured by the total number of reads aligned on the full-length sequence of the elements.

Read counts for genes and TEs were calculated separately for each sequenced library representing biological or technical replicates. Data counts were extracted with samtools (v. 1.3.1; Li et al., 2009) and were normalized using the EDAseq Bioconductor R-package (v. 2.18.0; Risso et al., 2011) in order to correct sequencing bias within each library (GC-content and gene length) and between libraries (sequencing depth) (Bullard et al., 2010; Li et al., 2015). These biases were corrected according to the full-quantile normalization implemented in EDASeq. Normalization factor with DESeq2 was skipped because differences in sequencing depth were already taken into account with EDASeq. After normalization, data quality assessment and sample homogeneity were analyzed by clustering (Euclidean distances between samples) and principal component analysis (PCA).

Statistical analyses of gene and TE expression were performed with the DESeq2 package (Love et al., 2014). This package calculates log2 fold-change between species (from normalized count), estimates the dispersion and uses Wald tests (negative binomial linear model) to determine significant differential expression. Before DESeq2 analyses, clusters of orthologs weakly expressed in all studied species (mean expression of all replicates lower than 5 reads after normalization step) were removed because (i) they potentially represented noise from RNA sequencing and (ii) their low counts could affect the statistical tests (biased dispersions and variance in log2 fold change). Regarding the limited number of biological replicates per species we considered for each species both biological and technical replicates without distinction, in order to maximize the intraspecific variance. Considering the large number of tests, adjusted p-values were estimated according to Bonferroni correction to minimize false-positive rates. In our analyses, genes and TEs were considered as differentially expressed (DE) between two species when the adjusted *p*-value was lower than 0.01.

Several comparisons were performed: (i) expression levels between species presenting the same ploidy or of the same hybrid origin (*S. versicolor* vs. *S. bakeri*; *S. alterniflora* vs.

S. maritima; *S.* × *townsendii* vs. *S.* × *neyrautii*), (ii) expression levels between the two tetraploid species (*S. versicolor* vs. *S. bakeri*) and the two hexaploid species (*S. alterniflora* vs. *S. maritima*), (iii) expression levels between each F1 hybrid (*S.* × *townsendii* vs. *S.* × *neyrautii*) and the MPV (Mid Parent Value; expected under additive parental expression) and (iv) expression levels between *S.* × *townsendii* and *S. anglica*. The MPV was calculated by averaging the expression levels found in both parental species (mean of expression levels found in all parental replicates after normalization). Genes that deviate from parental additivity in hybrids and allopolyploid (statistically DE compared to the estimated MPV) were classified in different patterns: under parental dominance when the gene expression level found in the hybrids or allopolyploid was not statistically DE compared to the expression level found in the maternal or paternal species (maternal or paternal dominance, respectively); transgressively expressed when the gene expression level found in the hybrids or allopolyploid was statistically DE compared to the expression levels found in maternal and in paternal species (transgressively up-regulated or down-regulated).

GO Enrichment Analysis

Gene ontology analyses were conducted to identify metabolic pathways or biological processes that were found differentially expressed following recent hybridization and genome doubling or between tetraploid and hexaploid species. GO term enrichment analyses were performed with the online platform agriGO (v. 2.0; Tian et al., 2017), considering the GO terms associated with *A. thaliana* (GO annotations available on TAIR10). The set of reference genes included only a subset of Arabidopsis genes for which homologous genes were found in *Spartina* transcriptomes. Biological processes over-represented in DE genes were determined using hypergeometric statistical tests using Singular Enrichment Analysis tool (Bonferroni multi-test adjustment, p-value < 0.01).

Parental Origin of Transgressively Expressed Orthologs

We selected a set of non-additive, transgressively expressed orthologs in the F1 hybrid *S.* × *townsendii* and/or the allopolyploid *S. anglica* compared to their parental species (*S. alterniflora* and *S. maritima*) and identified the parental origin of the transcribed hybrid/allopolyploid haplotypes (homeologous copies) involved in such an expression pattern. The detection of haplotypes was performed as previously described for high ploidy-level species without diploid genome reference (Boutte et al., 2016) by detecting polymorphisms (in regions with minimum read-depth of 30; with minimum proportion of reads displaying the alternative character state of 10%) and phasing reads sharing at least two character states in sliding windows of 120 bp.

RESULTS

New Reference Transcriptomes and Functional Annotations

The *de novo* transcriptome assembly process allowed building of new reference transcriptomes for the two tetraploid species containing 103,101 contigs for *S. versicolor* and 107,319 contigs for *S. bakeri*. Non-aligned reads to previously assembled transcriptomes of the hexaploid and allododecaploid species (40% of reads) were separately assembled, which resulted in the enrichment of reference transcriptomes from 158,825 to 240,710 contigs according to species (**Table 2**). Consequently, the new reference transcriptomes used for expression analyses contained 217,800 and 206,348 contigs for the parental hexaploid species *S. maritima* and *S. alterniflora*, respectively, 281,153 and 266,906 contigs in the F1 hybrids *S.* × *townsendii* and *S.* × *neyrautii*, and 297,327 contigs in the allododecaploid *S. anglica*. Their GC-contents range from 45 to 49% and their N50 varies from 417 bp in the F1 hybrids and the allopolyploid to 724 bp in the tetraploid species, with an intermediate value in the hexaploid parental species (471–494 bp). Quality of the new

TABLE 2 | *Spartina* reference transcriptomes and functional annotations.

Species	*S. maritima*	*S. alterniflora*	*S.* × *townsendii*	*S.* × *neyrautii*	*S. anglica*	*S. versicolor*	*S. bakeri*
Number of contigs in reference transcriptomes (Boutte et al., 2016)	58,975	43,521	58,120	62,101	56,617	–	–
Number of new assembled contigs	158,825	162,827	223,033	204,805	240,710	103,101	107,319
Total number of contigs in reference transcriptome	217,800	206,348	281,153	266,906	297,327	103,101	107,319
Number of annotated contigs	76,916 (35%)	69,980 (34%)	93,645 (33%)	91,798 (34%)	94,130 (32%)	66,924 (65%)	62,827 (59%)
Number of unigenes (detected using the *O. sativa* genome)	22,186	22,133	23,756	23,139	23,544	19,328	19,779
GC-content	45.18%	45.25%	46.10%	45.43%	45.33%	49.30%	47.23%
N50 (bp)	471	494	418	417	417	700	724
Mean contig length (bp)	345	351	313	322	315	559	565
Median contig length (bp)	230	226	211	223	209	358	356
Minimum contig length (bp)	100	100	100	100	100	100	100
Maximum contig length (bp)	10,313	10,839	9,423	14,705	8,833	17,112	16,705

reference transcriptomes was assessed using BUSCO with a panel of genes representing Poales species. Results are given in **Supplementary Figure 1**.

Functional annotations were assigned to 35 and 34% of contigs in *S. maritima* and *S. alterniflora*, respectively (**Table 2**). The vast majority of annotations was performed with the Blastx method (e.g., in *S. maritima*, Blastx allowed us to assign a function to 86% of annotated contigs and 48% with the HMM-Pfam method; **Supplementary Figure 2**). In comparison with parental species, the same proportions of contigs were also annotated in the F1 hybrids and the allopolyploid with 93,645 contigs (33%) in *S.* × *townsendii*, 91,798 contigs (34%) in *S.* × *neyrautii* and 94,130 contigs (32%) in *S. anglica*. Finally, within tetraploid transcriptomes, 65% of contigs in *S. versicolor* and 59% of contigs in *S. bakeri* were annotated (representing, respectively, 66,924 and 62,827 contigs). In order to reduce the computational time of analyses, we explored gene expression evolution for this set of annotated contigs, which represent a large part of the global gene expression in *Spartina* leaves (83% of mapped reads for each library **Supplementary Table 2**).

Orthologous Contig Identification

Orthologous contigs among reference transcriptomes were identified and 62,830 clusters of orthologs were formed using OrthoVenn2. It is important to note that a few clusters may sometimes belong to a single gene if it could not be fully assembled. Among these clusters, 12,520 contained at least one ortholog of each of the seven *Spartina* species, 1,725 contained only orthologs of hexaploid species and their derived taxa (orthologs found in transcriptomes of the parental species, the F1 hybrids and the allododecaploid) and 2,640 contained only orthologs of tetraploid species (**Figure 2**). In addition, 4,856 clusters contained exclusively orthologs identified in both F1 hybrids, and 2,399 orthologs of only *S.* × *townsendii* and *S. anglica*. Overall, the identification of orthologs with the OrthoVenn2 method allowed finding orthologs for more than 62% of annotated contigs [from 62% in *S. versicolor* (41,832 contigs) to 66% in *S. alterniflora* (50,245 contigs); **Supplementary Table 2**]. The composition of each cluster of orthologs (contig name, annotation) is detailed in **Supplementary Table 3**.

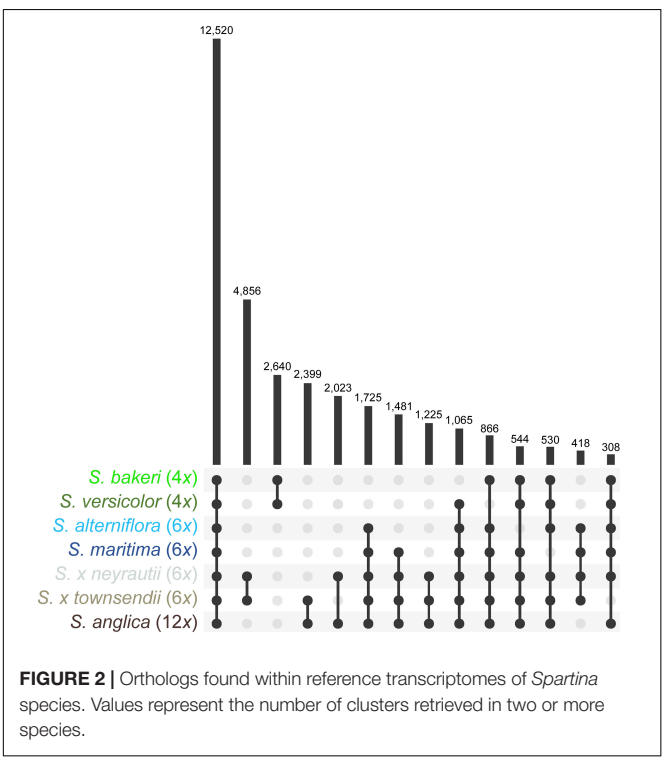


FIGURE 2 | Orthologs found within reference transcriptomes of *Spartina* species. Values represent the number of clusters retrieved in two or more species.

Gene Expression

After data normalization and elimination of low counts, 42,423 clusters of orthologs were retained for statistical analyses (representing 19,417 unigenes detected using the *O. sativa* genome). Data quality was assessed via hierarchical clustering (Euclidean distances) and principal component analysis of counts from each library (**Supplementary Figure 3**). Given the absence of aberrant results, all libraries from each species were conserved as replicates for analyses.

Expression Evolution in Tetraploid and Hexaploid Contexts

Comparisons between tetraploids showed that in leaves, 6,111 contigs (or orthologs) were significantly differentially expressed (DE) between *S. versicolor* and *S. bakeri*, representing 14.4% of the studied contigs (**Figure 3A**). Almost half of them were over-expressed in *S. bakeri* (3,178; 7.5%) the other half being over-expressed in *S. versicolor* (2,933; 6.9%). Orthologs more expressed in *S. bakeri* appear to be involved in cell development, modification and protein transport, whereas orthologs more expressed in *S. versicolor* are involved in epidermis development, response to biotic and abiotic stresses (salt, light, cold, and bacterium), gene silencing and post-transcriptional protein modifications (**Supplementary Table 4**).

More orthologs (15,278; 36.0%) were found DE between the hexaploid species *S. maritima* and *S. alterniflora* than between the tetraploids. Among them, 7,482 orthologs (48.9% of DE contigs) were over-expressed in *S. alterniflora* compared to *S. maritima* and were mainly involved in cell development, post-translational protein modifications (protein desumoylation, protein amino acid myristoylation), fatty acid and starch metabolic processes, response to salt and cold stresses and gene silencing (**Supplementary Table 4**). Conversely, 7,805 orthologs (51.1% of DE contigs) were over-expressed in *S. maritima* and were involved in chloroplast and epidermis development, protein catabolic process, post-translational protein modifications (amino acid phosphorylation and methylation), glycolysis process. When comparing tetraploids with hexaploids, the number of DE orthologs varies from 11,573 (27.3%) between *S. bakeri* and *S. alterniflora* to 14,911 (35.1%) between *S. versicolor* and *S. maritima* (**Figure 3A**). GO-term enrichment analyses showed that these DE orthologs have notably a key role in glycolysis process, epidermis development, pigment biosynthetic process and gene silencing (**Figure 3B** and **Supplementary Table 5**).

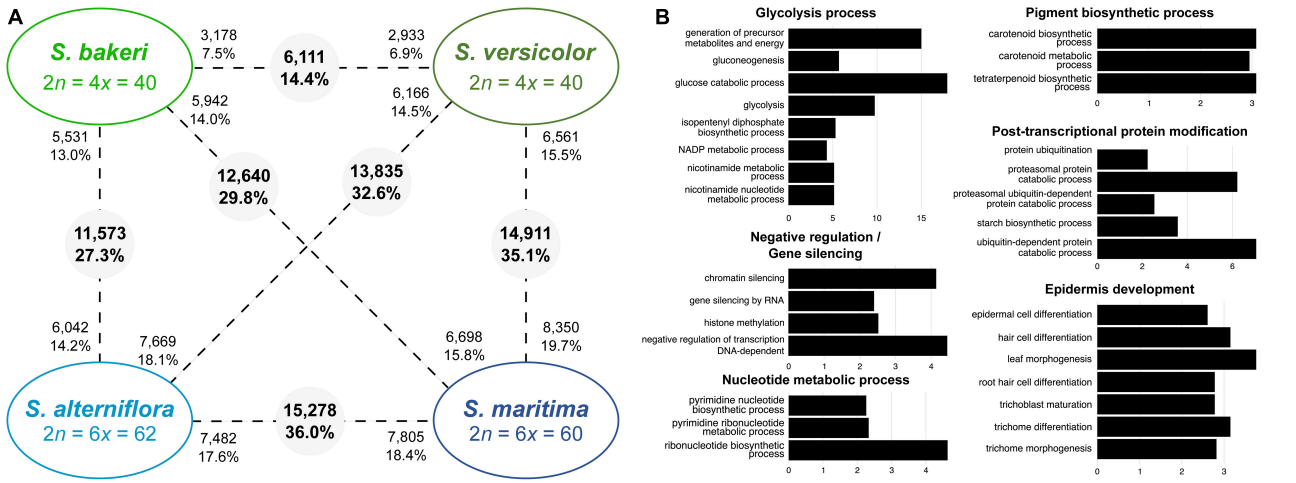


FIGURE 3 | Transcriptomic changes between tetraploids species (*S. bakeri* and *S. versicolor*) and hexaploid species (*S. alterniflora* and *S. maritima*). **(A)** Number of differentially expressed genes between species (pairwise comparisons). The numbers in grey circles represent the number of orthologs differentially expressed among the 42,423 analyzed clusters. The numbers of upregulated orthologs are represented near species names. **(B)** Bar graph of the most enriched GO-terms (biological process) for the differentially expressed orthologs between tetraploid and hexaploid species. GO-terms were grouped according to their metabolic pathway affiliation. *X*-axis corresponds to the negative logarithm of adjusted *p*-value. The complete list of enriched GO terms is available in **Supplementary Tables 4**, **5**.

Expression Evolution Following Recent Hybridization Events

In order to examine the transcriptomic effects of interspecific hybridization in the two independently formed F1 hybrids (with *S. alterniflora* as maternal parent and *S. maritima* as paternal parent), gene expression levels were compared first with the *in silico* MPV (average expression between parental species) (**Figure 4A**). Then orthologs that were found differentially expressed between the two hybrids were examined to (i) identify the potentially affected functions and (ii) in which way these DE genes deviate from parental additivity.

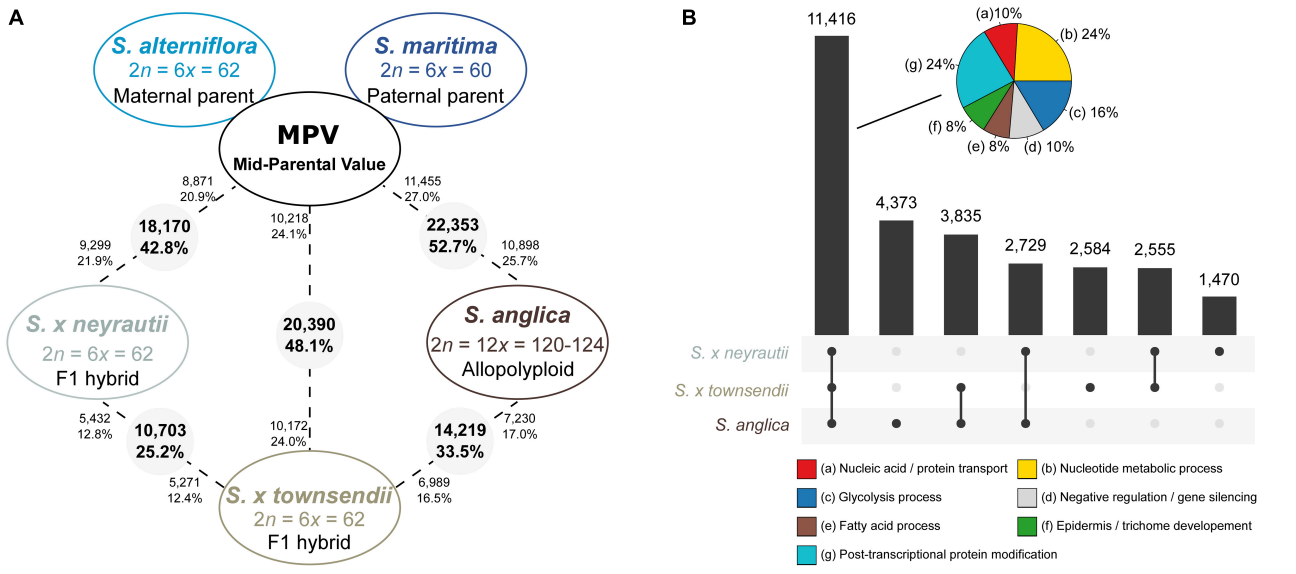


FIGURE 4 | Transcriptomic changes in F1 hybrids and the allopolyploid species compared to the *in silico* MPV (mid-parent value). **(A)** The number (and percentages) of differentially expressed (DE) genes between the MPV (expression expected under additive parental expression) and the F1 hybrids (*S.* × *townsendii* and *S.* × *neyrautii*) or the allopolyploid *S. anglica* are provided. The numbers in grey circles represent the number of orthologs differentially expressed among the 42,423 analyzed clusters. The numbers of upregulated orthologs are represented near species names. **(B)** Diagram of common DE genes between the MPV and species deriving from recent allopolyploidization. Pie charts represent enriched metabolic pathways within 11, 416 genes found DE compared to the MPV both among *S.* × *townsendii*, *S.* × *neyrautii*, and *S. anglica* (% of genes associated with enriched metabolic pathways). The complete list of enriched GO terms is available in **Supplementary Table 6**.

Comparisons of each hybrid with the MPV revealed that among 42,423 studied clusters of orthologs, 48.1% of clusters (20,390 orthologs) in *S.* × *townsendii* and 42.8% of clusters (18,170 orthologs) in *S.* × *neyrautii* deviated from parental additivity. For each hybrid, an equivalent number of DE orthologs was under-expressed compared to the MPV (10,218 and 8,871 orthologs, respectively in *S.* × *townsendii* and *S.* × *neyrautii*) and over-expressed compared to the MPV (10,172 and 9,299 orthologs, respectively in *S.* × *townsendii* and *S.* × *neyrautii*). Expression patterns in hybrids (**Table 3**) revealed that 70% of non-additive genes were expressed similarly to one of the parental species (parental expression dominance). In *S.* × *townsendii*, 6,908 orthologs (16.3%) mimicked the maternal parent (*S. alterniflora*) whereas 7,151 orthologs (16.9%) mimicked the paternal parent *S. maritima*. In *S.* × *neyrautii*, 6,421 orthologs (15.1%) mimicked the maternal parent *S. alterniflora* whereas 6,466 orthologs (15.2%) mimicked the paternal parent *S. maritima*. Other DE genes were found transgressively expressed in the hybrids compared to the MPV. Interestingly, in both hybrids a higher number of orthologs was more transgressively up-regulated (3,884 and 3,452 in *S.* × *townsendii* and *S.* × *neyrautii*, respectively) than transgressively down-regulated (2,447 and 1,831, respectively).

We found that among the genes deviating from parental additivity in hybrids (20,390 orthologs in *S.* × *townsendii* and 18,170 orthologs *S.* × *neyrautii* as indicated above), 13,971 of them were DE compared to MPV in both hybrids (**Figure 4B**). These genes were mainly involved in post-transcriptional protein modification processes, nucleotide metabolic processes and glycolysis (**Figure 4B** and **Supplementary Table 6**). Enriched GO-terms also belongs to gene silencing mechanisms (e.g., GO:0006342 chromatin silencing; GO:0016571 histone methylation; GO:0035196 production of miRNAs involved in gene silencing by miRNA). More specifically, 63% of DE orthologs (8,848/13,971) shared a similar pattern in both hybrids: 6,322 orthologs were under the same parental dominance and 2,526 orthologs were transgressively expressed in *S.* × *townsendii* and *S.* × *neyrautii* (**Table 4**). Enriched GO-terms for genes under paternal dominance in *S.* × *townsendii* and *S.* × *neyrautii* were mainly involved in gene silencing processes whereas genes under maternal dominance were involved in glycolysis process, post-transcriptional protein modification and protein transport (**Figure 5**). Other genes involved in the glycolysis process also appeared transgressively expressed in both F1 hybrids.

When comparing the two hybrids, we identified 10,703 DE orthologs (25.2% of clusters) between *S.* × *townsendii* and *S.* × *neyrautii* (**Figure 4A**). A large majority of these 25.2% orthologs (10,037; 93.8%) exhibits different expression patterns in each hybrid with regard to parental additivity (**Supplementary Table 7**). Specifically, 3,321 orthologs (31% of DE orthologs) were under parental dominance in both hybrids but did not mimic the same parent: 1,639 orthologs appeared under maternal dominance in *S.* × *townsendii* but under paternal dominance in *S.* × *neyrautii* and conversely, 1,682 orthologs appeared under paternal dominance in *S.* × *townsendii* and under maternal dominance in *S.* × *neyrautii*. These orthologs were involved in trichome development and response to stresses (salt, light, and cold) (**Supplementary Table 7**). Other DE orthologs between hybrids were mainly additive in one hybrid and expressed transgressively in the other hybrid (4,034 orthologs corresponding to 38% of DE orthologs). In total, 2,711 orthologs (25.3% of DE contigs) were transgressively up-regulated in one hybrid and additive in the other and 1,323 orthologs (12.4% of DE orthologs) were transgressively down-regulated in one hybrid and additive in the other (**Supplementary Table 7**). GO-term enrichment analyses reveal that orthologs transgressively up-regulated in *S.* × *townsendii* and additive in *S.* × *neyrautii* (1,378 orthologs) were involved in biotic/abiotic stress responses (virus,

TABLE 3 | Gene expression patterns in the two F1 hybrids and the allopolyploid compared to the parental species *S. alterniflora* (the maternal parent) and *S. maritima* (the paternal parent).

	No DE from the MPV	DE from MPV										
	Additivity			Maternal expression dominance		Paternal expression dominance		Transgressive down-regulation			Transgressive up-regulation		
	♀–H–♂	♀–H–♂	♀–H–♂	♀–H	♂–H	H–♂	♀–H	♀–H	♂–H	♀–H	H	H	H
S. × *townsendii*	21,129	422	482	3,028	3,880	3,260	3,891	1,842	317	288	3,404	256	224
F1 hybrid		22,033		6,908		7,151			2,447			3,884	
		(51.9%)		(16.3%)		(16.9%)			(5.8%)			(9.2%)	
S. × *neyrautii*	23,488	340	425	2,860	3,561	2,987	3,479	1,323	273	235	3,113	199	140
F1 hybrid		24,253		6,421		6,466			1,831			3,452	
		(57.2%)		(15.1%)		(15.2%)			(4.3%)			(8.1%)	
S. anglica	18,621	667	782	3,300	4,540	3,022	4,057	2,186	335	337	3,946	361	269
allopolyploid		20,070		7,840		7,079			2,858			4,576	
		(47.3%)		(18.5%)		(16,7%)			(6.7%)			(10.8%)	

For each taxon, the first line indicates the number of contigs from each pattern, the second line the total number of contig from each category and the third line indicates the percentage of contigs from each category.

bacterium, salt, and cadmium) and in gene silencing (production of small RNA involved in gene silencing by RNA). In comparison, orthologs transgressively up-regulated in *S.* × *neyrautii* and additive in *S.* × *townsendii* (1,333 orthologs) were involved in gravitropism, cellular development and also in responses to abiotic stresses (salt and cadmium).

Expression Evolution Following Recent Allododecaploid Speciation

Gene expression in *S. anglica* was compared (i) with the MPV and (ii) following the genome duplication *per se* by comparing gene expression of *S. anglica* with *S.* × *townsendii*. Analyses showed that 22,353 orthologs (52.7% of clusters) deviated from parental additivity in the allopolyploid, which is more than what was found in *S.* × *townsendii* (20,390 DE orthologs, 48.1% from MPV) (**Figure 4A**). The comparison against MPV revealed that 68% of DE orthologs in *S. anglica* were also DE in *S.* × *townsendii* (**Figure 4B**). These genes were mainly involved in glycolysis and fatty acid process, post-transcriptional protein modification process, nucleotide metabolic process and gene silencing (**Figure 4B** and **Supplementary Table 6**). As reported for the F1 hybrids, the majority of identified DE orthologs (67%) were under parental expression dominance (70% in *S.* × *townsendii*; **Table 3**). However, in *S. anglica*, the number of genes under maternal dominance was superior to genes under paternal dominance (7,840 orthologs expressed similarly as *S. alterniflora* against 7,079 expressed similarly as *S. maritima*). This is in accordance with species-species comparisons (**Figure 6A**), which showed a smaller number of DE orthologs between *S. anglica* and *S. alterniflora* (15,409; 36.3%) than between *S. anglica* and *S. maritima* (16,143; 38.1%). In addition, gene expression patterns revealed an increase of transgressively expressed orthologs in *S. anglica* (7, 434) compared to both F1 hybrids (6,331 in *S.* × *townsendii* and 5,283 in *S.* × *neyrautii*), with a large majority of them (62%) being transgressively up-regulated.

We found that 56.7% of the studied orthologs in *S. anglica* conserved the same pattern as the F1 hybrid after genome doubling (**Table 5**). Among these, 35.02% of the analyzed orthologs were additive in *S.* × *townsendii* and conserved the additivity in the allopolyploid *S. anglica*. Interestingly, in the F1 hybrid and in the allopolyploid, genes mimicking the expression found in the maternal parent *S. alterniflora* (3,074 orthologs; 7.2%) were mainly involved in trichome development (**Table 5** and **Figure 6B**). Conversely, genes expressed as the paternal parent *S. maritima* (3,050 orthologs; 7.2%) were involved in gene silencing and down-regulation processes. Among genes expressed transgressively in *S.* × *townsendii* and *S. anglica*, those involved in epidermis development appear downregulated whereas those involved in starch metabolism appear upregulated.

When examining gene expression evolution following genome duplication *per se*, 14,219 orthologs (33.5%) were found DE between the F1 hybrid *S.* × *townsendii* and the allopolyploid *S. anglica* (**Figure 6A**). An equivalent number of genes were up-regulated in each species: 6,989 orthologs were over-expressed in *S.* × *townsendii* and 7,230 orthologs were over-expressed in *S. anglica*. Genes more expressed in *S.* × *townsendii* appeared involved in chloroplast development, sulfur biosynthetic process, glycolysis process and response to stress (water deprivation, bacterium, and fungus infection) whereas genes more expressed in *S. anglica* appeared involved in epidermis development, negative regulation/gene silencing, response to salt stress (**Supplementary Table 8**). Most of the expression changes affected (i) genes that were additively expressed in *S.* × *townsendii* and found under maternal dominance or transgressively up-regulated in *S. anglica* (5.0% and 4.9% of orthologs, respectively) or, (ii) genes that were under paternal dominance in *S.* × *townsendii* and found under maternal dominance in *S. anglica* (5.0%) (**Table 5**). Finally, more genes shifted toward maternal dominance in *S. anglica* (7,840 orthologs) than to paternal dominance (7,079 orthologs).

Detection of haplotypes was performed for a set of 58 orthologous contigs showing an expression level clearly different following hybridization and duplication (transgressively expressed in *S.* × *townsendii* and *S. anglica*) and with enriched

TABLE 4 | Gene expression pattern evolution in the two independently formed hybrids *S.* × *townsendii* and *S.* × *neyrautii*.

Expression pattern in *S.* × *townsendii*	Expression pattern in *S.* × *neyrautii*	Number of genes	% of genes
Additive	Additive	17,834	**42.0%**
	Maternal expression dominance	1,069	2.5%
	Paternal expression dominance	984	3.3%
	Transgressive down-regulation	638	1.5%
	Transgressive up-regulation	1,508	3.6%
Maternal expression dominance	Additive	1,730	4.1%
	Maternal expression dominance	3,045	**7.2%**
	Paternal expression dominance	1,826	4.3%
	Transgressive down-regulation	155	0.4%
	Transgressive up-regulation	152	0.4%
Paternal expression dominance	Additive	1,733	4.1%
	Maternal expression dominance	1,868	4.4%
	Paternal expression dominance	3,277	**7.7%**
	Transgressive down-regulation	163	0.4%
	Transgressive up-regulation	110	0.3%
Transgressive down-regulation	Additive	1,150	2.7%
	Maternal expression dominance	216	0.5%
	Paternal expression dominance	209	0.5%
	Transgressive down-regulation	858	**2.0%**
	Transgressive up-regulation	14	0.1%
Transgressive up-regulation	Additive	1,806	4.3%
	Maternal expression dominance	223	0.5%
	Paternal expression dominance	170	0.4%
	Transgressive down-regulation	17	0.1%
	Transgressive up-regulation	1,668	**3.9%**

The percentage of genes that conserved the same expression pattern in both hybrids are in bold.

functions. Among these 58 contigs, eight were too weakly expressed in the parents, preventing the identification of parental haplotypes in hybrids/allopolyploids; five were displaying no polymorphism within and between species; 28 were displaying an additive pattern (in terms of presence/absence) of parental haplotypes in the hybrids and the allopolyploid (but expressed in a transgressive manner); five were non-additive in *S.* × *neyrautii*, six in *S.* × *townsendii*; one in both hybrids; and five were non-additive in both F1 hybrids and the allopolyploid (three of them are outlined in **Table 6**). The gene encoding the subunit SSRP1 of the FACT complex (Facilitates Chromatin Transcription) was transgressively upregulated in *S.* × *townsendii* and *S. anglica*. For this gene, only the haplotype originating from *S. alterniflora* was detected in the F1 hybrid and the allopolyploid, together with haplotypes shared by the two parents (i.e., of "undetermined origin"). The two other genes involved in salt stress response, showed different contributions of the

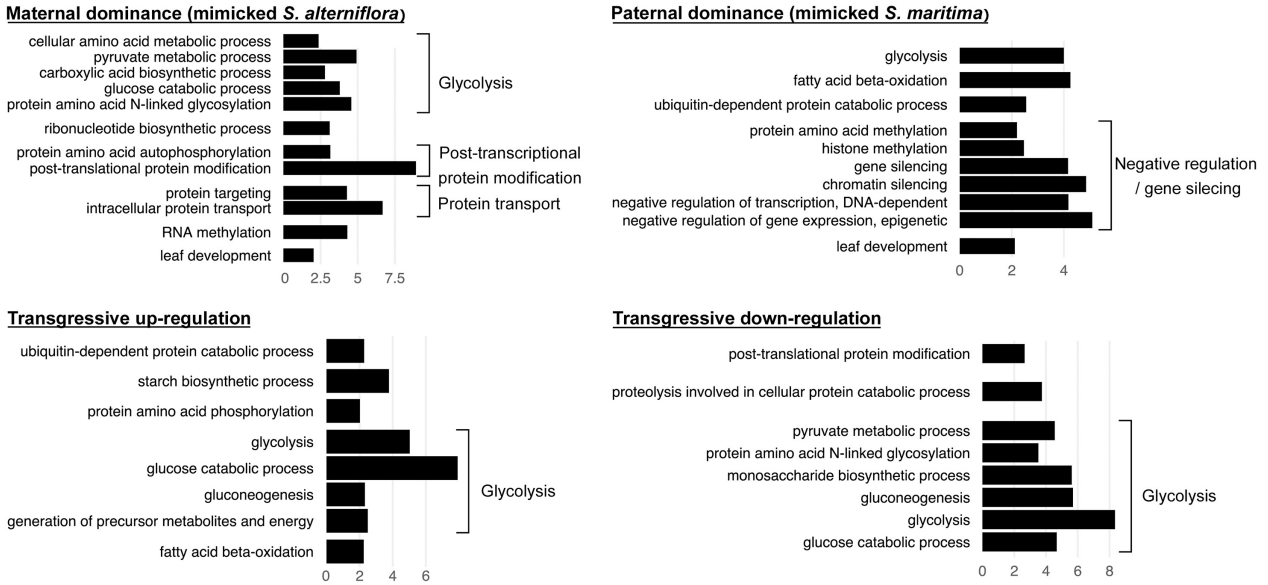


FIGURE 5 | Bar graph of the most enriched GO-terms (biological process) corresponding to genes found DE compared to the MPV and with the same expression patterns in both F1 hybrids *S.* × *townsendii* and *S.* × *neyrautii* (e.g., bar graph of maternal dominance represents the enriched functions of the 3,045 genes found under maternal dominance in *S.* × *townsendii* and in *S. neyrautii*). GO-terms were grouped according to their metabolic pathway affiliation. *X*-axis corresponds to the negative logarithm of adjusted *p*-value.

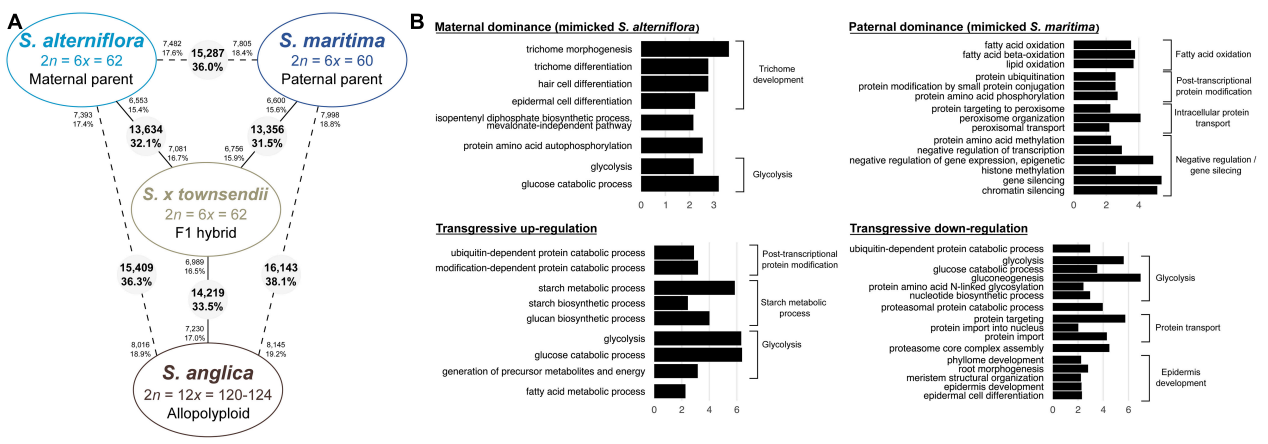


FIGURE 6 | Transcriptomic changes following hybridization and whole genome duplication in *Spartina*. **(A)** Number of differentially expressed genes between species (pairwise comparisons). The numbers in grey circles represent the number of orthologs differentially expressed among the 42,423 analyzed clusters. The numbers of upregulated orthologs are represented near species names. **(B)** Bar graph of the most enriched GO-terms (biological process) for non-additively expressed orthologs in the F1 hybrid and in the allopolyploid (expressed under parental dominance or transgressive). GO-terms were grouped according to their metabolic pathway affiliation. The *X*-axis corresponds to the negative logarithm of adjusted *p*-value. The complete list of enriched GO terms is available in **Supplementary Tables 6, 7**.

TABLE 5 | Evolution of gene expression patterns between the F1 hybrid *S.* × *townsendii* and the allopolyploid *S. anglica*.

Expression pattern in *S.* × *townsendii*	Expression pattern in *S. anglica*	Number of genes	% of genes
Additive	Additive	14,931	**35.2%**
	Maternal expression dominance	2,140	5.0%
	Paternal expression dominance	1,693	4.0%
	Transgressive down-regulation	1,208	2.8%
	Transgressive up-regulation	2,061	4.9%
Maternal expression dominance	Additive	1,384	3.3%
	Maternal expression dominance	3,074	**7.2%**
	Paternal expression dominance	1,896	4.5%
	Transgressive down-regulation	268	0.6%
	Transgressive up-regulation	286	0.7%
Paternal expression dominance	Additive	1,455	3.4%
	Maternal expression dominance	2,109	5.0%
	Paternal expression dominance	3,050	**7.2%**
	Transgressive down-regulation	268	0.6%
	Transgressive up-regulation	269	0.6%
Transgressive down-regulation	Additive	795	1.9%
	Maternal expression dominance	284	0.7%
	Paternal expression dominance	256	0.6%
	Transgressive down-regulation	1,081	**2.5%**
	Transgressive up-regulation	31	0.1%
Transgressive up-regulation	Additive	1,505	3.5%
	Maternal expression dominance	233	0.5%
	Paternal expression dominance	184	0.4%
	Transgressive down-regulation	33	0.1%
	Transgressive up-regulation	1,929	**4.5%**

The percentage of genes that conserved the same pattern of expression in both species are in bold.

TABLE 6 | Origin of haplotypes identified for three genes transgressively up-regulated following hybridization and genome doubling in *S.* × *townsendii* and *S. anglica.*

Haplotype origin	Haplotypes inherited from *S. alterniflora*	Haplotypes inherited from *S. maritima*	Undetermined parental origin (with no polymorphism)
Subunit SSRP1 of FACT complex			
Haplotypes identified in parental species	✔ (2)		✔ (1)
Origin of haplotypes identified in *S.* × *townsendii* (F1 hybrid)	✔ (2)		
Origin of haplotypes identified in *S. anglica* (Allopolyploid)	✔ (2)		
Chloride channel protein CLC-d			
Haplotypes identified in parental species	✔ (6)		✔ (5)
Origin of haplotypes identified in *S.* × *townsendii* (F1 hybrid)	✔ (3)		✔ (3)
Origin of haplotypes identified in *S. anglica* (Allopolyploid)	✔ (4)		✔ (4)
Salt-stress inducible tonoplast intrinsic protein 2			
Haplotypes identified in parental species	✔ (1)	✔ (2)	✔ (1)
Origin of haplotypes identified in *S.* × *townsendii* (F1 hybrid)		✔ (1)	
Origin of haplotypes identified in *S. anglica* (Allopolyploid)	✔ (1)	✔ (2)	

The number of haplotypes detected is indicated in brackets.

parental copies in their transgressive over-expression in the F1 hybrid and in the allopolyploid. For the CLC-d gene encoding a chloride channel protein, all haplotypes retrieved in the parental species were detected in *S.* × *townsendii* and *S. anglica*. However, for the gene encoding the salt-stress inducible tonoplast aquaporin 2, only haplotypes inherited from *S. maritima* were detected in *S.* × *townsendii*, whereas in *S. anglica*, haplotypes from both parental species were detected.

TE Expression

A strong transcriptional activity was detected for nine TE families in the analyzed *Spartina* species (**Figure 7A**). Most of them were Class I elements such as LTR-retrotransposons [2 *Gypsy* lineages (*Tekay* and *Ogre*) and four *Copia* lineages (*SIRE, Ivana, Ikeros,* and *Ale*), *LINE* and *SINE*]. Concerning Class II elements, only *CACTA* lineages were notably expressed. *LINE* elements appear as the most highly expressed TE (**Figure 7A**) with an average expression level two times higher than *Gypsy Tekay* (the second highest expressed TE) and four times higher than *Copia* lineages. Expression levels of other less expressed TEs were shown in **Supplementary Figure 4**. Among these, mutator elements are significantly more expressed in the tetraploid *S. bakeri* than in the other tetraploid or hexaploid species.

These nine TE families were transcriptionally active in all *Spartina* species but their expression levels varied according to species (**Figure 7B**). The two tetraploid species showed similar and moderate TE expression levels compared to the hexaploid species. Indeed, all *Copia* lineages, as well as the *Gypsy Tekay*, *LINE* and *CACTA*, were significantly less expressed in *S. versicolor* and *S. bakeri* than in *S. maritima* and/or in *S. alterniflora*. Only *SINE* elements were more expressed in both tetraploids than in hexaploids (twofold more expressed in tetraploids).

Expression comparisons between the parental hexaploid species *S. maritima* and *S. alterniflora* revealed the greatest number of differentially expressed TEs (six of the nine highest expressed TEs; **Figure 7B**) and the greatest variation in terms of expression levels. The *Gypsy Tekay*, Ogre, the *LINEs*

and the *CACTAs* elements were clearly more expressed in *S. maritima* than in *S. alterniflora* (twofold more expressed for *Gypsy Tekay*, *LINEs* and ninefold more for *CACTAs*) whereas the *Copia Ikeros* and Ale were more expressed in *S. alterniflora* than in *S. maritima* (eightfold and threefold more expressed, respectively).

In contrast, both F1 hybrids *S.* × *townsendii* and *S.* × *neyrautii* shared similar expression levels for all studied TEs. No significant differences were found except for the *Gypsy Tekay*, the *SINE* and *CACTA* elements (slightly more expressed in *S.* × *townsendii* than in *S.* × *neyrautii*). When comparing TE expression levels of the F1 hybrids with those of their parental species, statistical analyses demonstrated that most TEs were non-additively expressed in the F1 hybrids. Four of the nine studied TEs (**Figure 7B**) exhibited parental dominance in both or at least one F1 hybrid. The *Gypsy Tekay*, *Copia SIRE,* and *LINE* elements mimicked the expression of the maternal parent *S. alterniflora* whereas only *Gypsy Ogre* mimicked the expression of the paternal parent *S. maritima*. TEs under parental dominance in F1 hybrids displayed expressions similar to the parent with the lowest expression, especially when the parental species exhibited large expression differences. The same pattern was also observed for *SINE* elements that showed a transgressive downregulated expression in both F1 hybrids (slightly less expressed in *S.* × *townsendii* and *S.* × *neyrautii* than in the parental species).

Finally, analysis of TE expression of the allopolyploid *S. anglica* revealed that the down-regulation patterns observed in *S.* × *townsendii* were reinforced in *S. anglica*. Indeed, *Gypsy Tekay*, *Ogre*, *Copia Ivana*, *Ikeros* as well as *SINE* and *CACTA* elements were significantly less expressed in *S. anglica* than in *S.* × *townsendii*. Comparisons with expression levels detected in the parental species revealed that the expression dominance, biased toward the parent with the lowest expression level, was also strengthened in *S. anglica*: *Gypsy Tekay*, *Ogre* and *Copia Ivana* elements were found under maternal dominance in *S. anglica*, in addition to *Copia SIRE* and *LINE*s already found under the same dominance in *S.* × *townsendii*.

DISCUSSION

Transcriptomic changes affecting genes and transposable elements are major responses to hybridization and polyploidy. These changes that may occur immediately after (allo)polyploid speciation and persist over long-term of evolutionary time, as well as their consequences in terms of species adaptation and ecology, are timely but challenging questions. Taking advantage from the *Spartina* system where recurrent hybridization and polyploidy are well-documented in a well-understood phylogenetic context (Marchant, 1968b; Baumel et al., 2002a; Ainouche et al., 2009), we evaluated transcriptomic changes following past and recent polyploid speciation events by comparing gene and TE expressions among seven *Spartina* species with ploidy ranging from 4× to 12×. New reference transcriptomes were assembled for the tetraploid *S. bakeri* and *S. versicolor* (*syn. S. patens*), and the previously annotated transcriptomes were enriched for the hexaploid species (*S. alterniflora*, *S. maritima*, *S.* × *townsendii,* and *S.* × *neyrautii*) and the allododecaploid *S. anglica* (Ferreira de Carvalho et al., 2013b; Boutte et al., 2016). Ortholog detection and statistical analyses of gene and TE expression allowed comparing the expression levels of orthologous contigs included in 42,423 clusters and 21 TE lineages.

Contrasted Gene Expression and TE Dynamics in Tetraploid and Hexaploid *Spartina* Subclades

The tetraploid lineage ($2n = 4x = 40$) is composed of eight *Spartina* species native to North or South America that so far have been poorly investigated regarding transcriptomic analyses (except the prairie cordgrass *Spartina pectinata*; Gedye et al., 2010; Nah et al., 2016). Here we analyzed two additional tetraploid species, *S. bakeri* and *S. versicolor*. These species show different ecological preferences (freshwater habitat and high marsh, respectively) but are weakly divergent morphologically (Mobberley, 1956) and genetically (Baumel et al., 2002a, 2016; Fortune et al., 2007; Peterson et al., 2014). Our results also show a weak transcriptome divergence as only a limited number of genes (6,111 orthologs; 14.4% of clusters) were found differentially expressed between these species (compared with other studied *Spartina*, see below). Genes differentially expressed were mainly involved in conserved metabolic processes such as cell development, transport and post-transcriptional protein modifications. However, in *S. versicolor* we observed a higher expression of genes involved in epidermis development. This result is consistent with epidermis differences identified by Maricle et al. (2009) who showed the presence of silica cells and a higher cuticle thickness in *S. versicolor* compared to *S. bakeri*. Specific analyses of genes involved in cuticle and silica cell development are necessary to confirm this causal relationship. In addition, *S. versicolor* displayed a higher expression of genes involved in abiotic stress including salt, light and cold stress response. *S. versicolor* is adapted to salt marsh conditions (Casolo et al., 2015) whereas *S. bakeri* occupies freshwater habitats. Higher tolerance to salt stress, as well as the presence of rhizomes in *S. versicolor* (absent in *S. bakeri*, Mobberley, 1956), may explain its invasiveness and its greater worldwide distribution compared to *S. bakeri*.

The majority of studied TE lineages exhibited similar expression levels in the two tetraploids. Previous investigations of repetitive sequences in the *S. bakeri* and *S. versicolor* genomes (Giraud et al., 2020) indicated that these species share the same quantity and diversity of TEs. Similar TE expression patterns thus indicate that no new significant transcriptional and/or transposition activity occurred following their divergence <2 mya, which occurred relatively recently in the *Spartina* clade (Rousseau-Gueutin et al., 2015). One exception can be made for Class II Mutator elements that appear to be ninefold more expressed in *S. bakeri* than in *S. versicolor* (**Supplementary Figure 3**), suggesting their recent transcriptional activation and

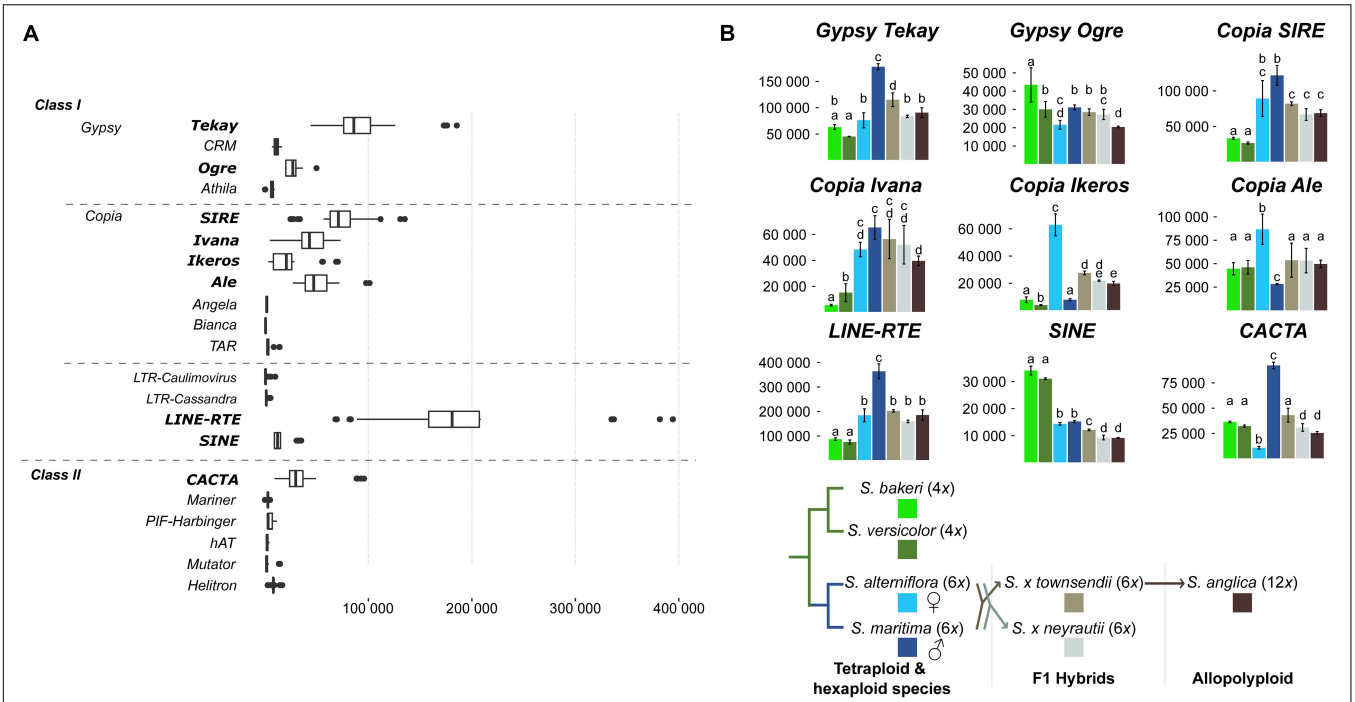


FIGURE 7 | Expression of transposable elements. **(A)** Expression levels comparison between all TEs identified in the *Spartina* genomes (Giraud et al., 2020). Boxplots were calculated from TE expression levels (normalized reads count) considering each library of all studied species. **(B)** Variation of TE expression levels between species. Only the nine highest expressed TEs are shown (p-value < 0.01). Letters above bars (from **a**–**e**) indicate significant differences of TE expression between species.

insertion in the *S. bakeri* genome. Quantitative estimations of *Mutator* elements indicate that they are notably more abundant in *S. bakeri* (9.4 Mb) than in *S. versicolor* (1.5 Mb) genomes (Giraud et al., 2020.), which would be consistent with their reported high rates of transposition (Dupeyron et al., 2019).

In contrast to the two tetraploids investigated, the hexaploids *S. maritima* ($2n = 6x = 60$) and *S. alterniflora* ($2n = 6x = 62$) are morphologically (Mobberley, 1956) and genetically (Baumel et al., 2001, 2002a; Salmon et al., 2005; Parisod et al., 2009) well-differentiated species. They diverged about 2–4 mya along the European and American Atlantic coasts, respectively (Rousseau-Gueutin et al., 2015). About 1–5% nuclear nucleotide divergence was reported between these species (Fortune et al., 2007; Chelaifa et al., 2010a; Ferreira de Carvalho et al., 2013a,b; Boutte et al., 2016). Genome wide expression patterns are consistent with this divergence, as 15,278 differentially expressed orthologs (36.0% of studied clusters) were detected between *S. alterniflora* and *S. maritima*. Genes DE were involved in several conserved metabolic processes (cell development, protein catabolic process, glycolysis process, transport, and post-transcriptional protein modifications), but interestingly also in processes related to the physiology and ecology of *Spartina*. As previously observed using quantitative PCR analyses on target genes (Ferreira de Carvalho et al., 2017), our study confirmed that genes involved in response to salt stress were up-regulated in *S. alterniflora* compared to *S. maritima*. This result was directly in line with better tolerance to abiotic stress and invasiveness of *S. alterniflora* highlighted in several studies that investigated its leaf morphology and anatomy (Maricle et al., 2009) and its tolerance to salt and hydrocarbon stress conditions (Watts et al., 2006; Bedre et al., 2016; Cavé-Radet et al., 2019b). These results somehow differ from previous transcriptomic comparisons between *S. maritima* and *S. alterniflora* performed using rice microarrays (Chelaifa et al., 2010a), which identified only 13.3% of DE genes (1,247 among the 9,353 examined genes in both species), most of them (957, belonging to developmental and cellular growth genes) being upregulated in *S. alterniflora* and downregulated in *S. maritima*. Various causes may explain the observed differences between our results and Chelaifa et al. (2010b) findings, such as: (i) rice microarray specificity, which tends to reduce the number of analyzed genes (70% of rice genes were hybridized with *Spartina* RNA on microarray), (ii) the potentially variable expression patterns in different conditions (i.e., plants maintained in the Greenhouse in Chelaifa's study and plants in natural conditions in this study) and different tissues (i.e., leaves and roots in Chelaifa's study and leaves in this study).

Divergent evolution between *S. maritima* and *S. alterniflora* appears to have also resulted in significant TE expression divergence. *Gypsy Tekay*, *Ogre*, *LINEs*, and *CACTAs* elements were significantly more expressed in *S. maritima* than in *S. alterniflora*, whereas *Copia Ikeros* and *Ale* elements were more expressed in *S. alterniflora* than in *S. maritima*. The expression levels seem positively correlated with their relative

abundance in both genomes. For example, *Copia Ikeros* elements, which represent 25 Mb of the *S. alterniflora* genome and 8.9 Mb of the *S. maritima* genome (Giraud et al., 2020), were eightfold more expressed in *S. alterniflora* than in *S. maritima*. These results clearly show that TE activity in *S. alterniflora* and *S. maritima* has differently evolved since their divergence 2–4 mya and probably led to some TE insertions in one or the other species according to TE lineage. Cavé-Radet et al. (2019a) identified 3,730 ra-siRNAs involved in the TE regulation in *S. maritima* and *S. alterniflora* preferentially targeting *Copia Ivana* and *SIRE*, *Gypsy Tekay* and *LINE* elements. These findings as well as our expression analyses indicate that in both hexaploids, these four active TEs are post-transcriptionally regulated (via small RNA synthesis) preventing their accumulation in the genome. However, highly expressed TEs such as *Copia Ikeros* and *Ale* in *S. alterniflora* and *CACTAs* in *S. maritima*, were not clearly under smallRNA control (Cavé-Radet et al., 2019a). This would suggest that both *Copia* elements benefit from *trans*-regulation of other expressed *Copia* elements (such as *Ivana* and *SIRE*). The limited post-transcriptional regulation of *CACTA*s despite their high transcriptional level may thus explain their accumulation in *S. maritima* (50.7 Mb vs. 26.0 Mb in *S. alterniflora*; Giraud et al., 2020).

All-to-all comparisons between tetraploid and hexaploid species revealed that the number of DE genes ranged from 27.3% (between *S. bakeri* and *S. alterniflora*) to 35.1% (between *S. versicolor* and *S. maritima*). GO-terms enrichment analyses showed that differences between tetraploid and hexaploid mainly concern genes involved in glycolysis, post-transcriptional protein modification, epidermis development. Interestingly, genes involved in the biosynthesis of carotenoids were less expressed in tetraploids than in hexaploids. Studies on plant responses to zinc and phenanthrene stresses in *Spartina densiflora* or to chromium stress in *Spartina argentinensis* (syn. *S. spartinae*; $2n = 4x = 40$) exhibit a decrease of carotenoids in stress condition (Mateos-Naranjo et al., 2008; Redondo-Gomez et al., 2011; Redondo-Gómez et al., 2011). Observed variations of genes involved in the carotenoid metabolic pathway between tetraploid and hexaploid species may be due to stress conditions of salt marsh (salt stress, flooding, and pollution) or linked to functions of carotenoids in plants (antioxidant during photosynthesis, precursors for the abscisic acid synthesis). To date, it remains unclear whether transcriptomic changes observed between tetraploid and hexaploid species appeared (i) following ploidy increase, (ii) more progressively during *Spartina* evolution or (iii) both. The auto- or allo-polyploid origin of the tetraploid and hexaploid *Spartina* lineages is not yet fully elucidated, and divergence of the duplicated ancestral genomes (including regulatory elements) and their subsequent evolution must have affected the transcriptome fates of the studied species. Nuclear gene phylogenies or haplotype detection from RNA-Seq data performed so far have revealed the presence of differentiated homeologs in both tetraploid and hexaploids (Fortune et al., 2007; Boutte et al., 2016; Ferreira de Carvalho et al., 2017) which would suggest a reticulate (i.e., allopolyploid) origin of these lineages, although differentiating autopolyploidy (followed by duplicate gene divergence) from allopolyploidy in the meso-tetraploid species is a challenging task with no known related diploid species. Moreover, which ancestral tetraploid genome(s) contributed to the hexaploid ancestor of *S. maritima* and *S. alterniflora* remains an open question.

Rapid Transcriptome Evolution Following Interspecific Hybridization: Alteration of Gene Expression and TE Silencing

The *Spartina* system offers unique opportunities to explore two components of the allopolyploid speciation process in natural conditions (hybridization versus genome duplication), a situation rarely met among recent and natural allopolyploid models (Ainouche and Wendel, 2014). Moreover, the first steps of the allopolyploid speciation process in natural conditions, i.e., consequences of divergent genome merger, can be explored in two independently formed hybrids *S.* × *townsendii* and *S.* × *neyrautii,* which share the same maternal (*S. alterniflora*) and paternal (*S. maritima*) species (Baumel et al., 2003; Ainouche et al., 2004). These two hybrids exhibit different morphologies, *S.* × *townsendii* bearing intermediate traits between the parental species, and *S.* × *neyrautii* being highly similar to *S. alterniflora* (Marchant, 1977; Baumel et al., 2003). In spite of high pollen sterility, *S.* × *townsendii* still forms vigorous populations at the hybridizing site in England (Renny-Byfield et al., 2010; Huska et al., 2016), whereas only remnant sterile *S.* × *neyrautii* individuals are surviving in south-west France as a result of site disturbance and urbanization (Hubbard, 1968; Baumel et al., 2003). The different traits of these two hybrids sharing the same genetic origin have always been puzzling. Our analyses reveal that hybridization resulted in consistent transcriptomic changes, with slightly more genes deviating from parental expression additivity in *S.* × *townsendii* (48.1%) than in *S.* × *neyrautii* (42.8%). These results confirmed the gene expression alteration following hybridization previously found in *S.* × *townsendii* and *S.* × *neyrautii* by microarray analyses (Chelaifa, 2010; Chelaifa et al., 2010b). The majority of genes considered DE compared to MPV in *S.* × *townsendii* and *S.* × *neyrautii* were found DE in both hybrids indicating consistent effects of two independent hybridization events on gene regulation evolution. More specifically, in both *S.* × *townsendii* and *S.* × *neyrautii* genes under paternal dominance were mainly involved in gene silencing processes whereas genes under maternal dominance were involved in glycolysis process, post-transcriptional protein modification and protein transport. Our results also revealed that genes involved in negative regulation and chromatin silencing were overexpressed in *S. alterniflora* compared to *S. maritima*. This negative regulation mimicking the paternal parent in hybrids, not previously reported, is consistent with DNA methylation repatterning identified following hybridization using methylation sensitive amplified polymorphism analyses (Salmon et al., 2005; Parisod et al., 2009). Results thus suggest that epigenetic modifications, which appear rapidly in the newly formed F1 hybrids, led to gene repression or silencing in *S.* × *townsendii* and *S.* × *neyrautii.* Similar transcriptomic

differences were observed in rice hybrids and were shown to result from the differential production of small interfering RNA (siRNA) in the parental lines. Thus, the F1 hybrids produced siRNA at an intermediate level compared to the two parents and if the level of a siRNA produced was sufficient to methylate both homeologs, this led to the hypermethylation of the locus from the unmethylated parent and its lower transcription (Chodavarapu et al., 2012), as potentially observed in our *Spartina* hybrids.

In contrast with previous microarray analyses (Chelaifa, 2010), no significant parental expression dominance was observed for most genes in both F1 hybrids, with an equivalent number of genes under paternal and maternal dominance (16.9 vs. 16.3% in *S.* × *townsendii* and 15.2% vs. 15.1% in *S.* × *neyrautii*). However, our results agree with Chelaifa (2010) regarding genes differentially expressed between the two hybrids where over-expression of genes involved in development and cellular growth in *S.* × *neyrautii* compared to *S.* × *townsendii* was reported. Our study also found that genes transgressively up-regulated in *S.* × *neyrautii* and additive in *S.* × *townsendii* compared to MPV were mainly involved in gravitropism, development and cellular organization. These results are consistent with the known morphological differences between these hybrids: *S.* × *neyrautii* plants exhibit high stems, long and fleshy leaves and long rhizomes whereas *S.* × *townsendii* usually display small height with reduced leaves (Marchant, 1977). In agreement with the microarray study, we also found that genes involved in salt stress response were up-regulated in *S.* × *townsendii* compared to *S.* × *neyrautii*. Comparison of salt stress tolerance between these hybrids was not performed to date but could help to better understand their physiology, ecology and contrasted distribution. Finally, genes involved in gene silencing also appear more expressed in *S.* × *townsendii* than in *S.* × *neyrautii* suggesting that hybridization induced a gene expression repression more important in *S.* × *townsendii*. Interestingly, when examining the non-additive patterns (compared to the MPV) of the genes DE between the two hybrids, we surprisingly found contrasted patterns in *S.* × *townsendii* and *S.* × *neyrauti*. Thus, our results provide new insights regarding the consequences of divergent genome merger, where non-additive gene expression in independent hybridization events (involving similar parental genotypes, Baumel et al., 2003) may entail differential parental gene expression repatterning, which very well illustrates the myriads of possible outcomes resulting from the "genomic shock" of hybridization (Nieto Feliner et al., 2020). The observed expression changes could also reflect post-hybridization or post-genome duplication evolution, although this might be limited regarding the generation time in these perennial young (c.a. 150 years old) plants. Hybridization is widespread and recurrent in natural populations, which increases the hybrid population genetic diversity. Recurrent and reciprocal hybridization between diploid *Tragopogon dubius* and *T. pratensis* resulted in the formation of morphologically diverse neo-allotetraploid *T. miscellus* individuals, which exhibit consistent transcriptomic differences, departure from parental additivity and differential homeologous expression bias (Shan et al., 2020). In our case, *S.* × *neyrautii* and *S.* × *townsendii* share the same maternal parent (*S. alterniflora*, Baumel et al., 2003), in spite of morphological and transcriptomic differences, indicating different outcomes from the "replay of the evolution tape" (Gould, 1989).

Contrasted patterns are observed for repetitive sequences where TE transcriptional activity was similar in both F1 hybrids. Indeed, *Gypsy*, *Copia*, *LINEs*, *CACTAs* lineages expressed in the parental species were also highly expressed in *S.* × *townsendii* and *S.* × *neyrautii*. Both F1 hybrids exhibited similar levels and patterns of expression suggesting that both hybridization events induced similar consequences on TE activity. Comparisons with parental expression allowed classifying TEs into two main categories. On one hand, *Gypsy Ogre*, *Copia Ikeros* and *Ale* as well as *CACTA* elements were additively expressed in the F1 hybrids indicating no expression evolution following hybridization. On the other hand, *Gypsy Tekay*, *Copia SIRE*, *Ivana,* and *LINE* elements were repressed in both hybrids (mimicking the parental species displaying the lowest expression level). Small RNA analysis conducted by Cavé-Radet et al. (2019a) revealed that this second category of TEs was specially targeted by ra-siRNAs in hexaploid species. Decrease of their transcriptional activity thus seems directly assignable to small RNA regulation inherited from parental species. In addition, Parisod et al. (2009) showed, using methyl-sensitive transposon display, that following hybridization in *Spartina*, DNA methylation increased drastically near TEs. Small RNAs can also act upstream to induce DNA methylation (RdDM pathways; Axtell, 2013; Borges and Martienssen, 2015; Wendte and Schmitz, 2018) and thus reinforce TE silencing. Consequently, all these results suggest that activity of several TEs in *Spartina* hybrids was constrained by various epigenetic regulations established rapidly following hybridization, causing TE silencing and preventing putative TE burst. This is consistent with the "genomic quiescence" with no transposition burst reported for a few targeted transposable elements in *Spartina* hybrids by Baumel et al. (2002b) and Parisod et al. (2009).

Maternal Dominance and Strengthening of TE Silencing in the Neo-Allopolyploid

In the recently formed allododecaploid *S. anglica*, we analyzed the superimposed effects of hybridization and genome duplication on one hand, and the effect of genome duplication *per se* on the other hand, by comparing expression levels of the allododecaploid *S. anglica*, with those of the hexaploid parental species *S. alterniflora* and *S. maritima* and with those of *S.* × *townsendii*. About 33.5% contigs were DE between *S.* × *townsendii* and *S. anglica*. This result demonstrates that additional gene regulation changes affected gene expression after genome doubling *per se* as previously shown in *Spartina* (Chelaifa et al., 2010b), and in other allopolyploids such as *Senecio* (Hegarty et al., 2006), *Arabidopsis* (Wang et al., 2004; Madlung et al., 2005), *Triticum* (Shaked et al., 2001) or *Tragopogon* (Buggs et al., 2011; Boatwright et al., 2018). Among the main transcriptomic changes following genome duplication, it appeared that the number of genes mimicking the maternal expression pattern (*S. alterniflora*) increased in *S. anglica*

with 7,840 genes (18.5%) under maternal dominance. This is novel finding compared to previous analyses using microarrays (Chelaifa et al., 2010b) where the maternal dominance in *S.* × *townsendii* was found attenuated in *S. anglica*. Interestingly, genes involved in trichome development were under maternal dominance in *S.* × *townsendii* (mimicking *S. alterniflora*) and stayed under the same parental dominance in *S. anglica*. Several studies (e.g., in *Arabidopsis* and *Medicago sativa*; Alkio et al., 2005; Alves et al., 2017; Cavé-Radet et al., 2020) have shown the key role of trichomes in organic xenobiotics detoxification. In *Spartina*, tolerance to phenanthrene (PAH, polycyclic aromatic hydrocarbons) was increased following allopolyploidization (Cavé-Radet et al., 2019b). *S. alterniflora* was found more tolerant to phenanthrene than *S. maritima* but less tolerant than *S.* × *townsendii* and *S. anglica*. Then the trichome development regulation in the hybrid and allopolyploid, inherited from the maternal parent *S. alterniflora* may explain in part their better tolerance to PAH.

Conversely, in *S.* × *townsendii* and *S. anglica*, genes involved in silencing (chromatin silencing, histone methylation, and production of smallRNA) mimicked expression levels found in the paternal parent *S. maritima*. This result confirms that the transcriptomic repression observed following hybridization was inherited after genome doubling. This is consistent with a previous MSAP study that showed in *S. anglica* the inheritance of epigenetic marks, appearing following hybridization (Salmon et al., 2005; Parisod et al., 2009). Further comparisons between genes downregulated in *S.* × *townsendii* and *S. anglica* and miRNA-target genes identified in *Spartina* (Cavé-Radet et al., 2019a) will be interesting, to explore the link between miRNA production and decrease of such gene expression in hybrid and allopolyploid compared to parental species.

In addition to maternal dominance, an increase of transgressive genes was identified in *S. anglica* (7,434 genes in *S. anglica* vs. 6,331 in *S.* × *townsendii*), a phenomenon already observed by Chelaifa et al. (2010b). Genes transgressively expressed in the F1 hybrids and *S. anglica* (up or down-regulated compared to the parental lines) were involved in different biological processes, including epidermis development, starch metabolic process, post-transcriptional protein modifications. Among a selected set of 58 transgressively expressed genes, 28 of them showed that copies from both parental species contribute to the transgressive pattern. But in contrast, for 30 of them only one parental subgenome was involved in transgressive pattern.

Detecting homeologs in an allododecaploid species such as *Spartina anglica* is particularly challenging (Boutte et al., 2015, 2016), as the parental species are hexaploids, most likely of hybrid origins (Fortune et al., 2007) and as there is no known diploid *Spartina* that could be used as reference. The origins of the hexaploid lineage as well as the number of differentiated genomes that have been merged in the hexaploid ancestor are not fully elucidated yet. Notwithstanding this complexity and the absence of *Spartina* reference genome, we were able to take advantage from the divergence that occurred between the two hexaploid parents *S. maritima* and *S. alterniflora* in the last 2–4 my to identify polymorphic orthologous regions and detect *maritima* versus *alterniflora* haplotypes, using parental reference transcriptomes, a procedure successfully employed in recent allopolyploids [e.g., *Capsella bursa-pastoris,* Kryvokhyzha et al. (2019); *Mimulus peregrinus*, Edger et al. (2017); *Tragopogon mirus* and *T. miscellus*, Boatwright et al. (2018)]. Further analyses on the *Spartina* genomes, aiming at exploring the nature and history of monoploid genomes in tetraploids and hexaploids, will allow distinguishing the meso-homeologs in the modern polyploids.

Non-additive patterns of parental expression contribute to enhance plasticity and adaptive responses to fluctuating environments in natural allopolyploid populations (Ferreira de Carvalho et al., 2017; Shimizu-Inatsugi et al., 2017). Our results also show an over-expression of genes involved in salt stress response in *S. anglica* compared to the F1 hybrid and the parental species. Several studies reported the intercontinental invasiveness of *S. anglica* populations that cope with severe chemical or physical constraints on salt marshes and high salinity levels (An et al., 2007; Strong and Ayres, 2013; Ainouche and Gray, 2016; Wong et al., 2018). This increased tolerance in *S. anglica* could be explained by the up-regulation of genes involved in salt stress response after genome doubling. For example, haplotype detection on two up-regulated genes in *S. anglica* (compared to *S.* × *townsendii*) and specially involved in salt stress response showed two different ways of haplotype expression evolution. For the CLC-d gene encoding the chloride channel protein (Wang et al., 2015), increase of expression was induced by the up-regulation of all or only a part of parental haplotypes retrieved in *S. anglica*. However, for the gene encoding the salt-stress inducible tonoplast aquaporin 2 (Wang et al., 2014), the restoration of "*alterniflora*-type" haplotype expression may account for its up-regulation.

In addition, we observed an over-expression of genes involved in silencing in *S. anglica* compared to *S.* × *townsendii*. For example, the FACT gene is known to encode a histone chaperone that can mediate nucleosome disassembly and reassembly (Grasser, 2020). Modification of the chromatin states via such histone chaperone was shown to mediate gene expression programs and help plants to more efficiently cope with stressful conditions (Probst and Mittelsten Scheid, 2015). Thus, it is tempting to speculate that such a gene may have played a key role during the allopolyploidization process, which merged two different genomes with divergent regulatory elements in the same nucleus. The identification of the parental origin of haplotypes suggests that the up-regulation of this gene after genome doubling was due to the increase of the expression of copies from *S. alterniflora* (maternal subgenome) and silencing of "*maritima*-type" haplotypes (already observed in *S.* × *townsendii*). The upregulation of genes involved in silencing in *S. anglica* suggested not only a parental legacy of repression as indicated above but also a strengthening of repression after genome duplication.

In this paper, we compared the relative expression per transcriptomes in the hexaploid parents and F1 hybrids and

the allododecaploid, by normalizing the number of reads relative to the total transcripts in each species. This assumes that mRNA transcriptome size (total number of transcripts per cell) is constant, which holds true when cells produce similar levels of RNA/cell (Lovén et al., 2012). Although widely used in comparative RNA-Seq studies, this approach does not take into account the potential absolute mRNA transcriptome size variation (reviewed Coate and Doyle, 2015) that may result from increased ploidy, which affects both gene copy number and cell size for some tissues as reported when comparing the allotetraploid *Glycine dolichocarpa* to its diploid progenitors (Coate and Doyle, 2010). These authors developed a quantitative PCR (qPCR) assay that normalizes individual gene expression to the genomic copy number when comparing the allotetraploid *G. dolichocarpa* to its diploid progenitors. Relative mRNA transcriptome sizes could be estimated by coupling this essay with transcriptome-normalized expression data (RNA-Seq) and revealed that the allotetraploid leaf transcriptome was approximately 1.4-fold larger than either diploid progenitor transcriptome. In *Spartina anglica*, gene dosage response is more complex as the parental genomes are hexaploid, and their genome composition as well as the extent of the fractionation process has to be elucidated as mentioned above. Future studies paralleling co-extracted genomic DNA and RNA Illumina sequencing should allow exploring this question. A larger sampling of additional plant organs and tissues, including root and inflorescences would be of particular interest for exploring previously reported differential tolerance to stresses (i.e., HAP and salt-stress) of investigated species and with regard to their fertility. As we are aware that transcriptome changes do not obviously correlate to the translatome (e.g., Coate et al., 2014; Soltis et al., 2016), another interesting question would be the impact of these transcriptional changes on the proteome of *S. anglica*, and its correlation to transcriptome responses.

Regarding repetitive sequences, TE repression, already observed in F1 hybrids, seems reinforced following genome doubling in *S. anglica*. Among the nine highest expressed TEs, six of them (*Gypsy Tekay*, *Ogre*, *Copia Ivana*, *Ikeros*, *SINEs,* and *CACTAs*) were less expressed in *S. anglica* than in *S.* × *townsendii*. This indicates that additional regulation takes place following genome doubling to repress TE transcriptional activity. Studies on *Arabidopsis* or *Triticum* polyploids showed that allopolyploidization induced rapid methylation changes near TEs, avoiding TE burst (Madlung et al., 2005; Parisod et al., 2010a; Ben-David et al., 2013). Analyses in *Spartina* also revealed that DNA methylation that appeared near TEs following hybridization were conserved after genome doubling (Parisod et al., 2009). Therefore, in *S. anglica*, TE transcriptional activity was partly under the control of repressive epigenetic marks already implemented following hybridization and inherited from *S.* × *townsendii*.

In conclusion, our comparative transcriptomic analyses among *Spartina* species allowed us to understand the evolution of gene and TE expression following recent and past polyploidization events. Gene expression changes were consistent with phylogenetic relationships and divergence time between species. Comparisons of tetraploid and hexaploid species showed that the TE dynamics was clearly different, reflecting a complex evolutionary history in both lineages since their divergence 6–10 mya. Particularly remarkable is the significant transcriptome repatterning following reticulate evolution, where expression changes (consistent with epigenetic and regulatory mechanisms alterations) that took place in 150–170 years old hybrids and neo-allopolyploid *S. anglica* far exceeded long term divergent transcriptome evolution in the meso-tetraploid and meso-hexaploid lineages. The superimposed polyploidization events which took place in the *Spartina* clade during the last 10–12 my offered increased opportunity to partition parental expression. Recent allopolyploidy provided springboards for new regulatory and expression patterns that played a central role in the species traits and ecology, including abilities to colonize stressful and fluctuating environments on saltmarshes, as particularly well illustrated in the worldwide invasive allododecaploid *S. anglica*. The genomic and transcriptomic resources being developed on this system now open new perspectives to explore the deepest history of the parental species, the extent of fractionation affecting the ancestral tetraploid and hexaploid genomes, and the way this dynamic affects adaptation and invasiveness of the modern species.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available at the NCBI SRA database on Bioproject PRJNA657699. Data used for reference transcriptome assemblies by Boutte et al. (2016) are available on BioProject PRJNA338100 (www.ncbi.nlm.nih.gov/sra/?term=PRJNA338100).

AUTHOR CONTRIBUTIONS

MA, AS, and MR-G designed the experiments. OL contributed to molecular experiments for RNA seq. DG, AS, MA, and MR-G analyzed the results and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by University of Rennes 1 (grant allocation to DG), European Union Seventh Framework Program: project 'Genomerge' [FP7-CIG- 2013–2017; Grant No. 333709 to MR-G (transcriptome sequencing)].

ACKNOWLEDGMENTS

The following institutions are thanked for their support: University of Rennes 1, UMR 6553 Ecosystèmes, Biodiversité, Evolution (Ecobio), Observatoire des Sciences de l'Univers (Rennes), INEE CNRS. This work benefited from the technical

platforms PEM and Ecolex (UMR Ecobio), Genouest. T. Fontaine and R. Bodiguel (UMR Ecobio) are thanked for their assistance in maintaining the *Spartina* plant collection. The reviewers are thanked for helpful comments and suggestions on the earliest version of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.589160/full#supplementary-material

REFERENCES

Ainouche, M., Chelaifa, H., Ferreira, J., Bellot, S., Ainouche, A., and Salmon, A. (2012). "Polyploid evolution in *Spartina*: dealing with highly redundant hybrid genomes," in *Polyploidy and Genome Evolution*, eds P. S. Soltis and D. E. Soltis (Berlin: Springer), 225–243.

Ainouche, M., and Gray, A. (2016). Invasive *Spartina*: lessons and challenges. *Biol. Invasions* 18, 2119–2122. doi: 10.1007/s10530-016-1201-7

Ainouche, M. L., Baumel, A., and Salmon, A. (2004). *Spartina anglica* C. E. Hubbard: a natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biol. J. Linn. Soc.* 82, 475–484. doi: 10.1111/j.1095-8312.2004.00334.x

Ainouche, M. L., Fortune, P. M., Salmon, A., Parisod, C., Grandbastien, M.-A., Fukunaga, K., et al. (2009). Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biol. Invasions* 11, 1159–1173. doi: 10.1007/s10530-008-9383-2

Ainouche, M. L., and Wendel, J. F. (2014). "Polyploid speciation and genome evolution: lessons from recent allopolyploids," in *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life*, ed. P. Pontarotti (Cham: Springer International Publishing), 87–113.

Akhunova, A. R., Matniyazov, R. T., Liang, H., and Akhunov, E. D. (2010). Homoeolog-specific transcriptional bias in allopolyploid wheat. *BMC Genomics* 11:505. doi: 10.1186/1471-2164-11-505

Albertin, W., Brabant, P., Catrice, O., Eber, F., Jenczewski, E., Chèvre, A.-M., et al. (2005). Autopolyploidy in cabbage (*Brassica oleracea* L.) does not alter significantly the proteomes of green tissues. *Proteomics* 5, 2131–2139. doi: 10.1002/pmic.200401092

Alkio, M., Tabuchi, T. M., Wang, X., and Colón-Carmona, A. (2005). Stress responses to polycyclic aromatic hydrocarbons in *Arabidopsis* include growth inhibition and hypersensitive response-like symptoms. *J. Exp. Bot.* 56, 2983–2994. doi: 10.1093/jxb/eri295

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Alves, W. S., Manoel, E. A., Santos, N. S., Nunes, R. O., Domiciano, G. C., and Soares, M. R. (2017). Detection of polycyclic aromatic hydrocarbons (PAHs) in *Medicago sativa* L. by fluorescence microscopy. *Micron* 95, 23–30. doi: 10.1016/j.micron.2017.01.004

An, S. Q., Gu, B. H., Zhou, C. F., Wang, Z. S., Deng, Z. F., Zhi, Y. B., et al. (2007). *Spartina* invasion in China: implications for invasive species management and future research. *Weed Res.* 47, 183–191. doi: 10.1111/j.1365-3180.2007.00559.x

Axtell, M. J. (2013). Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64, 137–159. doi: 10.1146/annurev-arplant-050312-120043

Baumel, A., Ainouche, M. L., Bayer, R. J., Ainouche, A. K., and Misset, M. T. (2002a). Molecular phylogeny of hybridizing species from the genus *Spartina* Schreb. (Poaceae). *Mol. Phylogenet. Evol.* 22, 303–314. doi: 10.1006/mpev.2001.1064

Baumel, A., Ainouche, M. L., Kalendar, R., and Schulman, A. H. (2002b). Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). *Mol. Biol. Evol.* 19, 1218–1227. doi: 10.1093/oxfordjournals.molbev.a004182

Baumel, A., Ainouche, M. L., and Levasseur, J. E. (2001). Molecular investigations in populations of *Spartina anglica* C.E. Hubbard (Poaceae) invading coastal Brittany (France). *Mol. Ecol.* 10, 1689–1701. doi: 10.1046/j.1365-294X.2001.01299.x

Baumel, A., Ainouche, M. L., Misset, M. T., Gourret, J.-P., and Bayer, R. J. (2003). Genetic evidence for hybridization between the native *Spartina maritima* and the introduced *Spartina alterniflora* (Poaceae) in South-West France: *Spartina x neyrautii* re-examined. *Plant Syst. Evol.* 237, 87–97. doi: 10.1007/s00606-002-0251-8

Baumel, A., Rousseau-Gueutin, M., Sapienza-Bianchi, C., Gareil, A., Duong, N., Rousseau, H., et al. (2016). *Spartina versicolor* fabre: another case of *Spartina* trans-Atlantic introduction? *Biol. Invasions* 18, 2123–2135. doi: 10.1007/s10530-016-1128-z

Bedre, R., Mangu, V. R., Srivastava, S., Sanchez, L. E., and Baisakh, N. (2016). Transcriptome analysis of smooth cordgrass (*Spartina alterniflora* Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to salinity. *BMC Genomics* 17:657. doi: 10.1186/s12864-016-3017-3

Ben-David, S., Yaakov, B., and Kashkush, K. (2013). Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. *Plant J.* 76, 201–210. doi: 10.1111/tpj.12285

Bennetzen, J. L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 65, 505–530. doi: 10.1146/annurev-arplant-050213-035811

Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Boatwright, J. L., McIntyre, L. M., Morse, A. M., Chen, S., Yoo, M.-J., Koh, J., et al. (2018). A robust methodology for assessing differential homeolog contributions to the transcriptomes of allopolyploids. *Genetics* 210, 883–894. doi: 10.1534/genetics.118.301564

Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085

Bortolus, A., Adam, P., Adams, J. B., Ainouche, M. L., Ayres, D., Bertness, M. D., et al. (2019). Supporting *Spartina* : interdisciplinary perspective shows *Spartina* as a distinct solid genus. *Ecology* 100:e02863. doi: 10.1002/ecy.2863

Bottani, S., Zabet, N. R., Wendel, J. F., and Veitia, R. A. (2018). Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* 23, 393–402. doi: 10.1016/j.tplants.2018.01.002

Boutte, J., Aliaga, B., Lima, O., Ferreira, de Carvalho, J., Ainouche, A., et al. (2015). Haplotype detection from next-generation sequencing in high-ploidy-level species: 45S rDNA gene copies in the hexaploid *Spartina maritima*. *G3* 6, 29–40. doi: 10.1534/g3.115.023242

Boutte, J., Ferreira, de Carvalho, J., Rousseau-Gueutin, M., Poulain, J., Da Silva, C., et al. (2016). Reference transcriptomes and detection of duplicated copies in hexaploid and allododecaploid *Spartina species* (Poaceae). *Genome Biol. Evol.* 8, 3030–3044. doi: 10.1093/gbe/evw209

Buggs, R. J. A., Renny-Byfield, S., Chester, M., Jordon-Thaden, I. E., Viccini, L. F., Chamala, S., et al. (2012). Next-generation sequencing and genome evolution in allopolyploids. *Am. J. Bot.* 99, 372–382. doi: 10.3732/ajb.1100395

Buggs, R. J. A., Zhang, L., Miles, N., Tate, J. A., Gao, L., Wei, W., et al. (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* 21, 551–556. doi: 10.1016/j.cub.2011.02.016

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform.* 11:94. doi: 10.1186/1471-2105-11-94

Casolo, V., Tomasella, M., De Col, V., Braidot, E., Savi, T., and Nardini, A. (2015). Water relations of an invasive halophyte (*Spartina patens*): osmoregulation and ionic effects on xylem hydraulics. *Funct. Plant Biol.* 42, 264–273. doi: 10.1071/FP14172

Cavé-Radet, A., Giraud, D., Lima, O., El Amrani, A., Aïnouche, M., and Salmon, A. (2019a). Evolution of small RNA expression following hybridization and allopolyploidization: insights from *Spartina* species (Poaceae, Chloridoideae). *Plant Mol. Biol.* 102, 55–72. doi: 10.1007/s11103-019-00931-w

Cavé-Radet, A., Rabhi, M., Gouttefangeas, F., and El Amrani, A. (2020). Do specialized cells play a major role in organic xenobiotic detoxification in higher plants? *Front. Plant Sci.* 11:1037. doi: 10.3389/fpls.2020.01037

Cavé-Radet, A., Salmon, A., Lima, O., Ainouche, M. L., and El Amrani, A. (2019b). Increased tolerance to organic xenobiotics following recent allopolyploidy in *Spartina* (Poaceae). *Plant Sci.* 280, 143–154. doi: 10.1016/j.plantsci.2018.11.005

Céccoli, G., Ramos, J., Pilatti, V., Dellaferrera, I., Tivano, J. C., Taleisnik, E., et al. (2015). Salt glands in the Poaceae family and their relationship to salinity tolerance. *Bot. Rev.* 81, 162–178. doi: 10.1007/s12229-015-9153-7

Chelaifa, H. (2010). *Spéciation Allopolyploïde et Dynamique Fonctionnelle du Génome Chez les Spartines*. Dissertation thesis. Rennes: Université de Rennes 1.

Chelaifa, H., Mahé, F., and Ainouche, M. (2010a). Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Mol. Ecol.* 19, 2050–2063. doi: 10.1111/j.1365-294X.2010.04637.x

Chelaifa, H., Monnier, A., and Ainouche, M. (2010b). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (Poaceae). *New Phytol.* 186, 161–174. doi: 10.1111/j.1469-8137.2010.03179.x

Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* 12:35. doi: 10.1186/1471-2105-12-35

Chen, Z. J. (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

Chen, Z. J., and Ni, Z. (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28, 240–252. doi: 10.1002/bies.20374

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4, 258–268. doi: 10.1038/s41477-018-0136-7

Chodavarapu, R. K., Feng, S., Ding, B., Simon, S. A., Lopez, D., Jia, Y., et al. (2012). Transcriptome and methylome interactions in rice hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12040–12045. doi: 10.1073/pnas.1209297109

Coate, J. E., Bar, H., and Doyle, J. J. (2014). Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *Plant Cell* 26, 136–150. doi: 10.1105/tpc.113.119966

Coate, J. E., and Doyle, J. J. (2010). Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. *Genome Biol. Evol.* 2, 534–546. doi: 10.1093/gbe/evq038

Coate, J. E., and Doyle, J. J. (2015). Variation in transcriptome size: are we getting the message? *Chromosoma* 124, 27–43. doi: 10.1007/s00412-014-0496-3

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Combes, M.-C., Dereeper, A., Severac, D., Bertrand, B., and Lashermes, P. (2013). Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol.* 200, 251–260. doi: 10.1111/nph.12371

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

del Pozo, J. C., and Ramirez-Parra, E. (2014). Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. *Plant Cell Environ.* 37, 2722–2737. doi: 10.1111/pce.12344

Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Dupeyron, M., Singh, K. S., Bass, C., and Hayward, A. (2019). Evolution of *Mutator* transposable elements across eukaryotic diversity. *Mob. DNA* 10:12. doi: 10.1186/s13100-019-0153-8

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Fabre, M. E. (1849). Description d'une nouvelle espèce de *Spartina*, abondante sur une portion du littoral meéditerranéen. *Ann. Sci. Nat. Bot. Biol.* 3, 122–125.

Ferreira de Carvalho, J., Boutte, J., Bourdaud, P., Chelaifa, H., Ainouche, K., et al. (2017). Gene expression variation in natural populations of hexaploid and allododecaploid *Spartina species* (Poaceae). *Plant Syst. Evol.* 303, 1061–1079. doi: 10.1007/s00606-017-1446-3

Ferreira de Carvalho, J., Chelaifa, H., Boutte, J., Poulain, J., Couloux, A., et al. (2013a). Exploring the genome of the salt-marsh *Spartina maritima* (Poaceae, Chloridoideae) through BAC end sequence analysis. *Plant Mol. Biol.* 83, 591–606. doi: 10.1007/s11103-013-0111-7

Ferreira de Carvalho, J., Poulain, J., Da Silva, C., Wincker, P., Michon-Coudouel, S., et al. (2013b). Transcriptome *de novo* assembly from next-generation sequencing and comparative analyses in the hexaploid salt marsh species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Heredity* 110, 181–193. doi: 10.1038/hdy.2012.76

Finigan, P., Tanurdzic, M., and Martienssen, R. A. (2012). "Origins of novel phenotypic variation in polyploids," in *Polyploidy and Genome Evolution*, eds P. S. Soltis and D. E. Soltis (Berlin: Springer), 57–76.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, 222–230. doi: 10.1093/nar/gkt1223

Flagel, L., Udall, J., Nettleton, D., and 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, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

Fortune, P. M., Schierenbeck, K. A., Ainouche, A. K., Jacquemin, J., Wendel, J. F., and Ainouche, M. L. (2007). Evolutionary dynamics of *Waxy* and the origin of hexaploid *Spartina species* (Poaceae). *Mol. Phylogenet. Evol.* 43, 1040–1055. doi: 10.1016/j.ympev.2006.11.018

Foucaud, J. (1897). Un *Spartina* inédit. *Ann. Soc. Sci. Nat. Char. Inf.* 32, 220–222.

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

Gedye, K., Gonzalez-Hernandez, J., Ban, Y., Ge, X., Thimmapuram, J., Sun, F., et al. (2010). Investigation of the transcriptome of prairie cord grass, a new cellulosic biomass crop. *Plant Genome J.* 3:69. doi: 10.3835/plantgenome2010.06.0012

Giraud, D., Lima, O., Huteau, V., Coriton, O., Boutte, J., Kovarik, A., et al. (2020). Evolutionary dynamics of transposable elements and satellite DNAs in polyploid *Spartina species*. *Plant Sci.* 302:110671. doi: 10.1016/j.plantsci.2020.110671

Gould, S. J. (1989). *Wonderful Life: The Burgess Shale and* The Natur*e of History*. New York, NY: W. W. Norton & Company.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Gray, A. J., and Benham, P. E. M. (eds) (1990). *Spartina anglica: A Research Review*. London: HMSO.

Grasser, K. D. (2020). The FACT histone chaperone: tuning gene transcription in the chromatin context to modulate plant growth and development. *Front. Plant Sci.* 11:85. doi: 10.3389/fpls.2020.00085

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Groves, H., and Groves, J. (1880).*Spartina townsendii*. *Reports of the Botanical Society and Exchange Club of the British Isles* 1:37.

Ha, M., Kim, E.-D., and Chen, Z. J. (2009). Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2295–2300. doi: 10.1073/pnas.0807350106

Hammond, M. E. R., and Cooper, A. (eds) (2002). *Turning the Tide: The Eradication of Invasive Species?: Proceedings of the International Conference on Eradication of Island Invasives*. Gland: IUCN.

Hegarty, M. J., Barker, G. L., Wilson, I. D., Abbott, R. J., Edwards, K. J., and Hiscock, S. J. (2006). Transcriptome shock after interspecific hybridization

in *Senecio* is ameliorated by genome duplication. *Curr. Biol.* 16, 1652–1659. doi: 10.1016/j.cub.2006.06.071

Higgins, J., Magusin, A., Trick, M., Fraser, F., and Bancroft, I. (2012). Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics* 13:247. doi: 10.1186/1471-2164-13-247

Hollister, J. D., and Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19, 1419–1428. doi: 10.1101/gr.091678.109

Hu, G., and Wendel, J. F. (2019). *Cis – trans* controls and regulatory novelty accompanying allopolyploidization. *New Phytol.* 221, 1691–1700. doi: 10.1111/nph.15515

Hubbard, C. E. (1968). *Grasses*. London: Penguin Books.

Hubbard, J. C. E. (1965). *Spartina* marshes in southern England: VI. Pattern of invasion in Poole Harbour. *J. Ecol.* 53, 799–813. doi: 10.2307/2257637

Huska, D., Leitch, I. J., de Carvalho, J. F., Leitch, A. R., Salmon, A., Ainouche, M., et al. (2016). Persistence, dispersal and genetic evolution of recently formed *Spartina* homoploid hybrids and allopolyploids in Southern England. *Biol. Invasions* 18, 2137–2151. doi: 10.1007/s10530-015-0956-6

Jackson, S., and Chen, Z. J. (2010). Genomic and expression plasticity of polyploidy. *Curr. Opin. Plant Biol.* 13, 153–159. doi: 10.1016/j.pbi.2009.11.004

Kashkush, K., Feldman, M., and Levy, A. A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* 33, 102–106. doi: 10.1038/ng1063

Kryvokhyzha, D., Milesi, P., Duan, T., Orsucci, M., Wright, S. I., Glémin, S., et al. (2019). Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). *PLoS Genet.* 15:e1008131. doi: 10.1371/journal.pgen.1008131

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lerat, E., Casacuberta, J., Chaparro, C., and Vieira, C. (2019). On the importance to acknowledge transposable elements in epigenomic analyses. *Genes* 10:258. doi: 10.3390/genes10040258

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, P., Piao, Y., Shon, H. S., and Ryu, K. H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinform.* 16:347. doi: 10.1186/s12859-015-0778-7

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., et al. (2012). Revisiting global gene expression analysis. *Cell* 151, 476–482. doi: 10.1016/j.cell.2012.10.012

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268

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

Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., et al. (2005). Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* 41, 221–230. doi: 10.1111/j.1365-313X.2004.02297.x

Marchant, C. J. (1963). Corrected chromosome numbers for *Spartina* x *townsendii* and its parent species. *Nature* 199:929. doi: 10.1038/199929a0

Marchant, C. J. (1968a). Evolution in *Spartina* (Gramineae): II. Chromosomes, basic relationships and the problem of *S. x townsendii* agg. *J. Linn. Soc. Lond. Bot.* 60, 381–409. doi: 10.1111/j.1095-8339.1968.tb00096.x

Marchant, C. J. (1968b). Evolution in *Spartina* (Gramineae): III. Species chromosome numbers and their taxonomic significance. *J. Linn. Soc. Lond. Bot.* 60, 411–417. doi: 10.1111/j.1095-8339.1968.tb00097.x

Marchant, C. J. (1977). Hybrid characteristics in *Spartina x neyrautii* Fouc., a taxon rediscovered in northern Spain. *Bot. J. Linn. Soc.* 74, 289–296. doi: 10.1111/j.1095-8339.1977.tb01182.x

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959

Maricle, B. R., Koteyeva, N. K., Voznesenskaya, E. V., Thomasson, J. R., and Edwards, G. E. (2009). Diversity in leaf anatomy, and stomatal distribution and conductance, between salt marsh and freshwater species in the C4 genus *Spartina* (Poaceae). *New Phytol.* 184, 216–233. doi: 10.1111/j.1469-8137.2009.02903.x

Mateos-Naranjo, E., Redondo-Gómez, S., Cambrollé, J., Luque, T., and Figueroa, M. E. (2008). Growth and photosynthetic responses to zinc stress of an invasive cordgrass. *Spartina densiflora*. *Plant Biol.* 10, 754–762. doi: 10.1111/j.1438-8677.2008.00098.x

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

Mobberley, D. G. (1956). *Taxonomy and Distribution of the Genus Spartina*. Retrospective Theses and Dissertations. Ames. IL: IOWA State University.

Nah, G., Lee, M., Kim, D.-S., Rayburn, A. L., Voigt, T., and Lee, D. K. (2016). Transcriptome analysis of *Spartina pectinata* in response to freezing stress. *PLoS One* 11:e0152294. doi: 10.1371/journal.pone.0152294

Nieto Feliner, G., Casacuberta, J., and Wendel, J. F. (2020). Genomics of evolutionary novelty in hybrids and polyploids. *Front. Genet.* 11:792. doi: 10.3389/fgene.2020.00792

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer.

Paredes-Páliz, K. I., Mateos-Naranjo, E., Doukkali, B., Caviedes, M. A., Redondo-Gómez, S., Rodríguez-Llorente, I. D., et al. (2017). Modulation of *Spartina densiflora* plant growth and metal accumulation upon selective inoculation treatments: a comparison of gram negative and gram positive rhizobacteria. *Mar. Pollut. Bull.* 125, 77–85. doi: 10.1016/j.marpolbul.2017.07.072

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., et al. (2010a). Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* 186, 37–45. doi: 10.1111/j.1469-8137.2009.03096.x

Parisod, C., Holderegger, R., and Brochmann, C. (2010b). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi: 10.1111/j.1469-8137.2009.03142.x

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.-A., and Ainouche, M. (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol.* 184, 1003–1015. doi: 10.1111/j.1469-8137.2009.03029.x

Peterson, P. M., Romaschenko, K., Arrieta, Y. H., and Saarela, J. M. (2014). A molecular phylogeny and new subgeneric classification of *Sporobolus* (Poaceae: Chloridoideae: Sporobolinae). *Taxon* 63, 1212–1243. doi: 10.12705/636.19

Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q. T., et al. (2010). Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytol.* 186, 135–147. doi: 10.1111/j.1469-8137.2009.03140.x

Prieto, J., Cires, E., Sánchez Corominas, T., and Vázquez, V. (2011). Systematics and management of natural resources: the case of *Spartina species* on European shores. *Biologia* 66:1011. doi: 10.2478/s11756-011-0109-z

Probst, A. V., and Mittelsten Scheid, O. (2015). Stress-induced structural changes in plant chromatin. *Curr. Opin. Plant Biol.* 27, 8–16. doi: 10.1016/j.pbi.2015.05.011

Redondo-Gomez, S., Andrades-Moreno, L., Parra, R., Valera-Burgos, J., Real, M., Mateos-Naranjo, E., et al. (2011). *Spartina densiflora* demonstrates high tolerance to phenanthrene in soil and reduces it concentration. *Mar. Pollut. Bull.* 62:9. doi: 10.1016/j.marpolbul.2011.05.018

Redondo-Gómez, S., Mateos-Naranjo, E., Vecino-Bueno, I., and Feldman, S. R. (2011). Accumulation and tolerance characteristics of chromium in a cordgrass Cr-hyperaccumulator, *Spartina argentinensis*. *J. Hazard. Mater.* 185, 862–869. doi: 10.1016/j.jhazmat.2010.09.101

Renny-Byfield, S., Ainouche, M., Leitch, I. J., Lim, K. Y., Le Comber, S. C., and Leitch, A. R. (2010). Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin. *Ann. Bot.* 105, 527–533. doi: 10.1093/aob/mcq008

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinform.* 12:480. doi: 10.1186/1471-2105-12-480

Robertson, M., Schrey, A., Shayter, A., Moss, C. J., and Richards, C. (2017). Genetic and epigenetic variation in *Spartina alterniflora* following the

deepwater horizon oil spill. *Evol. Appl.* 10, 792–801. doi: 10.1111/eva.12482

Rousseau, H., Rousseau-Gueutin, M., Dauvergne, X., Boutte, J., Simon, G., Marnet, N., et al. (2017). Evolution of DMSP (dimethylsulfoniopropionate) biosynthesis pathway: origin and phylogenetic distribution in polyploid *Spartina* (Poaceae, Chloridoideae). *Mol. Phylogenet. Evol.* 114, 401–414. doi: 10.1016/j.ympev.2017.07.003

Rousseau-Gueutin, M., Bellot, S., Martin, G. E., Boutte, J., Chelaifa, H., Lima, O., et al. (2015). The chloroplast genome of the hexaploid *Spartina maritima* (Poaceae, Chloridoideae): comparative analyses and molecular dating. *Mol. Phylogenet. Evol.* 93, 5–16. doi: 10.1016/j.ympev.2015.06.013

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

Santos, F. C., Guyot, R., do Valle, C. B., Chiari, L., Techio, V. H., Heslop-Harrison, P., et al. (2015). Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. *Chromosome Res.* 23, 571–582. doi: 10.1007/s10577-015-9492-6

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4069–4074. doi: 10.1073/pnas.1101368108

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). "BUSCO: assessing genome assembly and annotation completeness," in *Gene Prediction Methods in Molecular Biology*, ed. M. Kollmar (New York, NY: Springer), 227–245.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13, 1749–1759.

Shan, S., Boatwright, J. L., Liu, X., Chanderbali, A. S., Fu, C., Soltis, P. S., et al. (2020). Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). *Front. Genet.* 11:888. doi: 10.3389/fgene.2020.00888

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., and Shimizu, K. K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* 26, 193–207. doi: 10.1111/mec.13738

Soltis, D. E., Misra, B. B., Shan, S., Chen, S., and Soltis, P. S. (2016). Polyploidy and the proteome. *Biochim. Biophys. Acta Proteins Proteomics* 1864, 896–907. doi: 10.1016/j.bbapap.2016.03.010

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003

Song, M. J., Potter, B. I., Doyle, J. J., and Coate, J. E. (2020). Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. *Plant Cell* 32, 1434–1448. doi: 10.1105/tpc.19.00832

Stebbins, G. L. (1971). *Chromosomal Evolution in Higher Plants*. London: Edward Arnold.

Strong, D. R., and Ayres, D. R. (2013). Ecological and evolutionary misadventures of *Spartina*. *Annu. Rev. Ecol. Evol. Syst.* 44, 389–410. doi: 10.1146/annurev-ecolsys-110512-135803

Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., et al. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467. doi: 10.1111/nph.13491

Tayalé, A., and Parisod, C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140, 79–96. doi: 10.1159/000351318

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, 122–129. doi: 10.1093/nar/gkx382

Ungerer, M. C., Strakosh, S. C., and Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16, 872–873. doi: 10.1016/j.cub.2006.09.020

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

Vicient, C. M., and Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* 120, 195–207. doi: 10.1093/aob/mcx078

Visger, C. J., Wong, G. K.-S., Zhang, Y., Soltis, P. S., and Soltis, D. E. (2019). Divergent gene expression levels between diploid and autotetraploid *Tolmiea* relative to the total transcriptome, the cell, and biomass. *Am. J. Bot.* 106, 280–291. doi: 10.1002/ajb2.1239

Wang, J., Tian, L., Madlung, A., Lee, H.-S., Chen, M., Lee, J. J., et al. (2004). Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 167, 1961–1973. doi: 10.1534/genetics.104.027896

Wang, L.-L., Chen, A.-P., Zhong, N.-Q., Liu, N., Wu, X.-M., Wang, F., et al. (2014). The *Thellungiella salsuginea* tonoplast aquaporin TsTIP1;2 functions in protection against multiple abiotic stresses. *Plant Cell Physiol.* 55, 148–161. doi: 10.1093/pcp/pct166

Wang, S., Su, S. Z., Wu, Y., Li, S. P., Shan, X. H., Liu, H. K., et al. (2015). Overexpression of maize chloride channel gene ZmCLC-d in *Arabidopsis thaliana* improved its stress resistance. *Biol. Plant* 59, 55–64. doi: 10.1007/s10535-014-0468-8

Watts, A. W., Ballestero, T. P., and Gardner, K. H. (2006). Uptake of polycyclic aromatic hydrocarbons (PAHs) in salt marsh plants *Spartina alterniflora* grown in contaminated sediments. *Chemosphere* 62, 1253–1260. doi: 10.1016/j.chemosphere.2005.07.006

Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Wendte, J. M., and Schmitz, R. J. (2018). Specifications of targeting heterochromatin modifications in plants. *Mol. Plant* 11, 381–387. doi: 10.1016/j.molp.2017.10.002

Wong, J. X. W., Costantini, F., Merloni, N., Savelli, L., Geelen, D., and Airoldi, L. (2018). The widespread and overlooked replacement of *Spartina maritima* by non-indigenous *S. anglica* and *S. townsendii* in north-western Adriatic saltmarshes. *Biol. Invasions* 20, 1687–1702. doi: 10.1007/s10530-017-1654-3

Wu, J., Lin, L., Xu, M., Chen, P., Liu, D., Sun, Q., et al. (2018). Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genomics* 19:586. doi: 10.1186/s12864-018-4966-5

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, 52–58. doi: 10.1093/nar/gkz333

Yaakov, B., and Kashkush, K. (2012). Mobilization of *Stowaway*-like MITEs in newly formed allohexaploid wheat species. *Plant Mol. Biol.* 80, 419–427. doi: 10.1007/s11103-012-9957-3

Yoo, M.-J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Yoo, M.-J., Szadkowski, E., and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

Zhang, J., Liu, Y., Xia, E.-H., Yao, Q.-Y., Liu, X.-D., and Gao, L.-Z. (2015). Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7022–7029. doi: 10.1073/pnas.1515170112

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

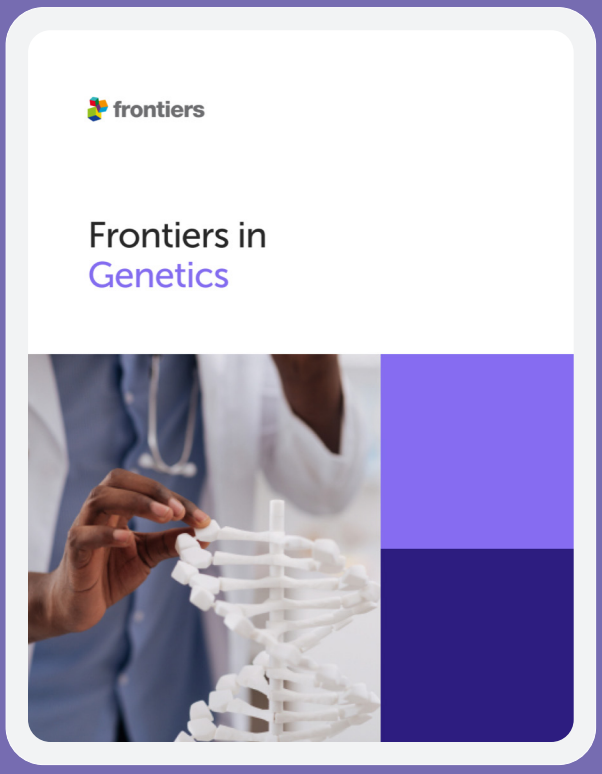
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact