ORIGINAL ARTICLE

Extensive changes in gene expression and alternative splicing due to homoeologous exchange in rice segmental allopolyploids

Zhibin Zhang[1] · Tiansi Fu[1] · Zhijian Liu[1] · Xutong Wang[2] · Hongwei Xun[1] · Guo Li[1] · Baoxu Ding[1] · Yuzhu Dong[1] · Xiuyun Lin[3] · Karen A. Sanguinet[4] · Bao Liu[1] · Ying Wu[1] · Lei Gong[1]

Abstract

Key message **We report rampant homoeologous exchanges in progenies of a newly synthesized rice segmental allo-tetraploid and demonstrate their consequences to changes of gene expression and alternative splicing.**

Abstract Allopolyploidization is recurrent across the tree of angiosperms and known as a driving evolutionary force in both plants and animals. A salient feature of allopolyploidization is the induction of homoeologous exchange (HE) events between the constituent subgenomes, which may in turn cause changes in gene expression, transcript alternative splicing, and phenotypic novelty. However, this issue has been poorly studied, largely because lack of a system in which the exact parentage donating the subgenomes is known and the HE events are occurring in real time. Here, we employed whole-genome re-sequencing and RNA-seq-based transcriptome profiling in four randomly chosen progeny individuals (at the 10th-selfed generation) of segmental allotetraploids that were constructed by colchicine-mediated whole-genome doubling of F1 hybrids between the two subspecies (*japonica* and *indica*) of Asian cultivated *Oryza sativa*. We show that rampant HE events occurred in these tetraploid individuals, which converted most of the otherwise heterozygous genomic regions into a homogenized state of one parental subgenome. We demonstrate that genes within these homogenized genomic regions in the tetraploids showed high frequencies of altered expression and enhanced alternative splicing relative to their counterparts in the corresponding diploid parents in the embryo tissue. Intriguingly, limited overlaps between the differentially expressed genes and the differential alternative spliced genes were identified, which were partitioned to distinctly enriched gene ontology terms. Together, our results indicate that HE is a major mechanism to rapidly generate novelty in gene expression and transcriptome diversity, which may facilitate phenotypic innovation in nascent allopolyploids and relevant to allopolyploid crop breeding.

Zhibin Zhang and Tiansi Fu authors contributed equally to this work.

✉ Ying Wu
  wuy003@nenu.edu.cn

✉ Lei Gong
  gongl100@nenu.edu.cn

Extended author information available on the last page of the article

Introduction

Polyploidization or whole-genome duplication (WGD) is a prominent driving force in evolution and speciation of eukaryotic organisms, especially in higher plants (Renny-Byfield and Wendel 2014; Soltis et al. 2015; Van de Peer et al. 2017). All flowering plants have experienced at least one round of polyploidization in their evolutionary history (Jiao et al. 2011). Based on the formation mode, polyploidy can be categorized into two major types, autopolyploids (WGD of one species' genome) and allopolyploids (WGD of merged genomes from two or more species) (Stebbins 1947; Doyle et al. 2003). However, under natural settings, this distinction can be blurred, with many intermediates between the two extremes. Thus, the term segmental allopolyploid refers to polyploidy formed between genetically distinct parental populations (e.g., subspecies) but within a given species (Stebbins 1947). In allo- and segmental allopolyploids,

concomitant merging and doubling of related but divergent genomes in a common nucleus and cytoplasm often catalyze chromosomal changes (Szadkowski et al. 2011; Zhang et al. 2013), accompanied with genomic changes and epigenetic modifications at the molecular level (Kraitshtein et al. 2010; Madlung and Wendel 2013; Chalhoub et al. 2014; Guo and Han 2014; Song and Chen 2015).

Among the various alterations in allopolyploids, homoeologous recombination appears ubiquitous (Gaeta and Pires 2010). Thus, chimeric or mosaic homogenized chromosomal regions from respective parents could result from nested and accumulated homoeologous recombination, causing replacement of homoeologous regions from one parent and elimination of the other (Gaeta and Pires 2010). This phenomenon has been termed homoeologous exchange (HE) (Salmon et al. 2010). HEs, as large-scale structural chromosome variations, occur frequently in both natural and synthesized allopolyploids (Flagel et al. 2012; Kovarik et al. 2012; Henry et al. 2014; Lashermes et al. 2014; Soltis et al. 2016; Lloyd et al. 2017). Most HE-related studies were performed in *Brassica* species which contain several allotetraploid crops, such as oilseed rape and mustard rape (He et al. 2016). HEs have been shown to cause aberrant meiotic behavior (Gaeta and Pires 2010), novel genetic variation such as gene presence/absence variations (PAVs) (Lashermes et al. 2014; Hurgobin et al. 2017) and are causal for many important phenotypic changes in *Brassica* polyploids (Zhao et al. 2006; Liu et al. 2012; Chalhoub et al. 2014). Recently, it has been reported that HEs can sustain and amplify the changes of DNA methylation and gene expression that can be decoupled from hybridity in earlier generations of the rice segmental allotetraploids (Li et al. 2019). Collectively, the importance of HEs in generating rapid genetic and phenotypic diversity, regulating gene expression, promoting adaptive genomic evolution and speciation of allopolyploids, has been increasingly recognized (Gaeta and Pires 2010; Szadkowski et al. 2011; Henry et al. 2014).

Analyses of gene expression alterations due to allopolyploidy have been focused on total gene expression and/or homoeologous expression, including non-additive gene expression, homoeologous expression bias, expression level dominance by using natural or newly synthetic allopolyploid systems (Buggs et al. 2011; Grover et al. 2012; Xu et al. 2014). However, transcriptomic consequence induced by HEs in allopolyploids is still poorly understood. Lloyd and colleagues have found that HEs caused extensive dosage-dependent gene expression changes in *B. napus* and also proposed the effects of HEs could be mitigated over time (Lloyd et al. 2017).

Compared with the known homogenized genomic landscape caused by HEs in allopolyploids, the issue of whether there are transcriptional expression changes of genes residing at the homogenized genomic regions relative to their homologous counterparts in the corresponding diploid parents is not yet investigated. It is largely due to the lack of appropriate allopolyploidy systems, in which the exact parentage donating the subgenomes is known and the HEs are newly occurred or still ongoing.

Alternative splicing (AS) is an important post-transcriptional process for regulating gene expression and generates transcriptome and proteome diversity in eukaryotes (Sultan et al. 2008; Syed et al. 2012; Kornblihtt et al. 2013). AS events are classified into six major types, including intron retention (IR), exon skipping (ES), alternative donor (AltD), alternative acceptor (AltA), alternative first exon (AltFE) and alternative last exon (AltLE) (Sturgill et al. 2013). In plants, AS presents many remarkable cases of functional consequences in developmental processes and environmental fitness, such as regulation of vernalization-mediated flowering (Rosloski et al. 2013; Marquardt et al. 2014), normal functioning of the circadian clock (Seo et al. 2012; Filichkin et al. 2015) and biotic immune and abiotic stress responses (Xu et al. 2012; Feng et al. 2014; Ling et al. 2015; Liu et al. 2017). Furthermore, AS alterations induced by polyploidization are an important aspect of adaptation and evolution since such events could generate immediate and dramatic changes in stoichiometry and activity of splicing factors which in turn regulate AS profiling at the whole-genome level (Syed et al. 2012). Accordingly, AS changes have been widely studied in natural and synthetic polyploids recently (Zhou et al. 2011; Saminathan et al. 2015; Wang et al. 2018b). For instance, by using RT-PCR and sequencing of 82 AS events, Zhou and colleagues found that many homoeologous pairs showed different AS patterns in natural allotetraploid *B. napus* and high proportion of AS event loss was identified in resynthesized allotetraploids compared to the parents (Zhou et al. 2011). Conversely, researchers have reported that the level of AS increased in tetraploid watermelon vegetative tissues compared with the corresponding diploid parent (Saminathan et al. 2015). A recent study uncovered that more than half of homoeologous genes produced divergent transcriptional isoforms in each subgenome of natural allotetraploid cotton (Wang et al. 2018b). However, ploidy-dependent AS regulation at the whole-genome level in allopolyploids remains scarcely studied, especially for genes mapped to the HEs-mediated homogenized genomic regions relative to their homologous counterparts in the respective parents.

Here, we employed whole-genome re-sequencing and an RNA-seq platform to trace the expression and alternative splicing patterns in embryonic tissue of four randomly chosen individual progeny of a segmental rice allotetraploid (at the 10th-selfed generation, S_{10}) constructed by colchicine-mediated WGD of an F1 hybrid between the two subspecies, *japonica* (cv. Nipponbare) and *indica* (cv. 9311), of Asian cultivated rice (*Oryza sativa* L.). In the present study, we

found that extensive HEs have homogenized homoeologous genomic regions. In addition to detecting extensive variation in the transcriptome profile induced by segmental allopolyploidization, we also found novel transcriptional changes (differentially expressed genes) and differential alternative splicing (AS) established within the homogenized genomic regions relative to their homologs in the respective parents. Finally, our results also implicate independent regulation of transcriptional gene expression and post-transcriptional AS in response to HEs within rice segmental allotetraploids. Taken together, our results suggest that HE represents a major mechanism to rapidly generate novelty in genomic composition, gene expression and transcript diversity, and hence, phenotypic innovation in nascent allopolyploid plants.

Materials and methods

Plant materials and growth conditions

F1 hybrids were produced by crossing two standard laboratory cultivars, Nipponbare (as the maternal parent) and 9311 (as the paternal parent), that represent the two subspecies, *japonica* and *indica*, respectively, of Asian cultivated rice (*Oryza sativa* L.). The segmental allotetraploid rice was generated by colchicine treatment on tillers of the F1 hybrid and was then self-pollinated consecutively for 10 generations (Xu et al. 2014). Four randomly selected 10th-selfed generation (S_{10})-old tetraploids were used in this study, together with the two parental rice cultivars and the F1 hybrid. All plant materials were grown in a standard long-day greenhouse conditions.

Sample preparation and sequencing

Leaves were harvested when the tetraploid seedlings reached the four-leaf stage and were then stored at -20 °C. Genomic DNA of four rice tetraploids was isolated from leaf tissue mentioned above by using a modified CTAB method (Allen et al. 2006) and phenol extractions. DNA quality was determined by a ND-2000 NanoDrop spectrophotometer (ThermoFisher Scientific, Inc., USA). The four tetraploid rice genomes were sequenced independently using the Illumina HiSeq 2500 platform. The 150 bp paired-end method was used for library construction with 300 bp as the average insert size, and for each sample, 43 gigabase (Gb) clean reads were obtained.

For each genotype (Nipponbare, 9311, F1 hybrid and four tetraploids), embryos were collected at 15 days after pollination (DAP), immediately frozen in liquid nitrogen and stored at -80 °C until used. Two biological replicates of independent individuals were used for all the diploid plant materials

(Nipponbare, 9311 and F1 hybrid); thus 10 samples in total were used in the RNA-seq analysis. Total RNA of each sample was extracted using TRIzol reagent (Invitrogen) based on the manufacturer's protocol. The 150 bp paired-end RNA-seq libraries with average insert size of 300 bp were constructed with TruSeq RNA Sample Preparation Kit v2 and sequenced on the Illumina HiSeq 2500 platform, and 112 gigabase (Gb) clean reads have been obtained for each sample.

All the clean data included both re-sequencing data and RNA-seq data have been deposited at the SRA database http://www.ncbi.nlm.nih.gov/sra/ under accession numbers PRJNA540689.

Re-sequencing data processing and HE regions determining

The clean data of each rice tetraploid DNA were mapped against the MSU7.0 Nipponbare genome sequence (http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/) by the BWA program (-n 5) (Li and Durbin 2009). A list of SNPs between each of the four tetraploids and Nipponbare genome DNA sequences were generated by the pileup module of Samtools (Li et al. 2009). Here, we also employed a SNPs list between 9311 and Nipponbare genomes we generated previously according to the simulated 9311 DNA sequencing data as the SNPs reference (hereafter referred to as SNPs-ref) (Xu et al. 2014). For each tetraploid, only the SNPs located in the SNPs-ref were marked as the reliable SNPs in the tetraploidy and were kept for further analysis, and all the genomic regions that didn't host SNPs that existed in SNPs-ref were Nipponbare-homogenized regions. Based on the reliable SNPs information of each tetraploid, the reads showing maternal and paternal origins within each 10 kb genomic window were counted (only the reads with > 5 depth were used), respectively, by using a serious of custom Perl scripts. Considering the parentage and the frequent homoeologous chromosome recombination events in the segmental rice tetraploids, there were only five possible parental original reads ratio/types of parental original genomic composition (i.e., Nipponbare:9311) for a given locus, and they were 0:4, 1:3, 2:2, 3:1 and 4:0. Then, for each 10 kb genomic window, the differences in parental original read ratio between each tetraploid (an actual ratio) and each of the five theoretical ratios above was detected by binomial distribution test followed by FDR corrections, and the exact criteria for identifying homogenized and heterozygous regions were listed below: (1) windows showing the proportion of Nipponbare-original reads > 90% were defined as Nipponbare-homogenized windows (or NPB-HG, for short); (2) windows with the proportion of Nipponbare-original reads < 10% were defined as 9311-homogenized windows

(or 9311-HG, for short); (3) windows with q-value > 0.05 compared with theoretical ratios of 1:3, 2:2 or 3:1 were accepted as the heterozygous windows; (4) windows showing different parental original genomic composition type compared with their adjacent 5 windows (i.e., 50 kb genomic region) were removed, and then the adjacent windows were merged to form large homogenized or heterozygous regions by Perl scripts if they showed the identical type of parental original genomic composition. Thereafter, homogenized and heterozygous genes were identified according to their genomic locations.

Identification of conservative gene models between Nipponbare and 9311

To obtain shared gene models between Nipponbare and 9311, the 9311 genome sequence was obtained from RISe (Rice Information System, http://rice.genomics.org.cn) and was used to simulate 50 million 150 bp paired-end genomic reads by using DWGSIM (http://davetang.org/wiki/tiki-index.php?page=DWGSIM). All simulated 9311 genomic DNA reads were mapped against the MSU7.0 Nipponbare genome by the BWA program (-n 5). Bedtools (Quinlan and Hall 2010) was used to calculate the coverage of 9311 reads to the exons with flanking upstream and downstream 50 bp regions of each Nipponbare reference non-transposon-related gene model. Only the gene models with $> 95\%$ consistency/coverage of 9311 DNA reads to the reference gene models from Nipponbare were reserved for subsequent gene expression and alternative splicing analysis.

Gene expression analysis

Clean RNA-Seq reads from each library were aligned to the MSU7.0 Nipponbare rice genome reference using TopHat v2.0.11 (segment mismatches 1, read-mismatches 2, max-multihits 20 and -r 0) (Trapnell et al. 2014). Proper aligned reads were counted by HTSeq software (Anders et al. 2015) for each sample. Normalization of library size for each sample was performed by DESeq v1.30.0 (Anders and Huber 2010). Only the homogenized genes specific for each tetraploid were reserved for further analysis, and based on their genomic compositions and locations, the homogenized genes were categorized into two groups of NPB-HG and 9311-HG genes. For a given tetraploid, differentially expressed homogenized genes between the tetraploid and each of the three diploids (Nipponbare, 9311 and F1 hybrid) were detected using DESeq v1.30.0 (Anders and Huber 2010), respectively. Genes showing significant differential expression level between tetraploids and the corresponding parent (i.e., FDR-adjusted p value (q value) < 0.05 with fold change > 2 when NPB-HG genes compared with the counterparts in Nipponbare and 9311-HG genes compared

with the counterparts of 9311) were defined as differentially expressed genes (DEGs) between the tetraploids and the corresponding parent, and the remaining genes were termed corresponding-parent-like genes (or cor-like genes, for short). Then, the DEGs were divided into two categories: (*a*) other-diploid-like DEGs (other-like DEGs, for short), those showed non-differential expression level compared with those in at least one of the alternative parents and F1 hybrid (i.e., q value ≥ 0.05 or q value < 0.05 but fold change ≤ 2 in tetraploids vs. the alternative parent or q value ≥ 0.05 or q value < 0.05 but fold change ≤ 2 in tetraploids vs. F1 hybrid); and (*b*) Tetraploid-novelty DEGs (tetra-novelty DEGs, for short) that showed significant differential expression between tetraploids and both the alternative parent and F1 hybrid (i.e., q value < 0.05 with fold change > 2 in tetraploids vs. the alternative parent and q value < 0.05 with fold change > 2 in tetraploids vs. F1 hybrid).

Annotation reconstruction and alternative splicing (AS) analysis

Complete transcript structures from each individual were assembled separately and all predicted transcript isoforms were merged using a Stringtie (Pertea et al. 2016) pipeline without any gene model annotations. The de novo annotation file (GTF format) was then compared with the MSU7.0 Nipponbare rice genome annotation file (GFF3 file from ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/) using Cuffcompare (Trapnell et al. 2014). Next, only the isoforms with the class code information of "j" and "=" were kept to generate the reconstructed annotation information. Finally, the expression levels of each isoform in all samples were calculated and tabulated according to the reconstructed annotation and isoforms which possessed an FPKM value > 1 in at least one sample that were reserved to format the final curated annotation file.

The *Spankijunc* module in the Spanki pipeline (Sturgill et al. 2013) and ASTALAVISTA program (Foissac and Sammeth 2007) were used to detect confident splicing junctions and AS events, respectively, as previously reported with some modifications (Wang et al. 2016). Only splice junctions with corresponding RPM (Reads per Million) value > 0.1 and entropy value > 2 were kept for downstream analysis. After merging the information for splice junctions from all samples, each AS event involved corresponding AS junctions was quantified with the definition of PSI (Percent Spliced Index = inclusion reads/(inclusion reads + exclusion reads)) using the *Spankisplicing* module and reliable AS events were defined based on the criteria that the number of reads supporting junction in inclusion path ≥ 5 and the PSI value > 0.05. To confirm that the identified IR (Intron Retention) events were not resulted by sequencing

of immature pre-mRNAs, *Transdecoder* (Haas et al. 2013) was utilized to estimate and compare the ratios of novel transcripts harboring complete ORFs in IR versus that of all background novel transcripts hosting other AS events (Wang et al. 2018a). Ultimately, reliable AS events were compared between tetraploids and three diploids (Nipponbare, 9311 and F1 hybrid) using the *Splicingcomp* module. DAS (differential alternative splicing) events between two compared samples were defined when the FDR-corrected q value of Fisher's exact test with correction was smaller than 0.05 and the absolute value of $\Delta PSI > 0.1$ (Wang et al. 2016). All steps of intermediate conversion of files during the AS analyses were performed with custom Perl scripts.

Similar to DEG, DAS (differential alternative splicing) was also defined. Two groups of DAS events were then defined as other-diploid-like DAS (other-like DAS, for short) and tetraploid-novelty DAS (tetra-novelty DAS), based on the q-value of Fisher's exact test with correction and ΔPSI value. Other-like DAS events met the criteria that: (i) $\Delta PSI > 0.1$ and q value < 0.05 between AS event belonged to homogenized regions in the tetraploid and the counterpart in the corresponding parent and (ii) $\Delta PSI < 0.1$ or FDR-corrected q value < 0.05 between AS events belonged to homogenized regions in the tetraploid and the counterpart in the alternative parent or F1 hybrid, whereas tetra-novelty DAS events met the criteria that $\Delta PSI > 0.1$ and FDR-corrected q value < 0.05 between AS events belonging to homogenized regions in the tetraploid and the counterparts in all three diploids. Genes harboring any DAS events were defined as DSGs (DAS-related genes).

Gene ontology (GO) analysis for both DEG and DSG

Blast2GO v2.5.0 (https://www.blast2go.com/) was performed to assign gene ontology (GO) terms. Further, annotation of the genes for GO mapping was restricted by GO Slim assignment. GO terms with FDR-corrected q value < 0.05 were considered significantly overrepresented by hypergeometric tests in the R packages of ClusterProfiler (Yu et al. 2012).

Validation of DAS events using reverse-transcription (RT) PCR analysis

The total embryonic RNA samples of all tested plant materials were treated with DNaseI (Invitrogen), reverse-transcribed by the SuperScript-TM-RNase H-Reverse Transcriptase (Invitrogen), and subjected to RT-PCR analysis using gene-specific primers. We randomly selected 10 DAS events (six events for IR and four events for ES, respectively) to verify the DAS events happened between the tetraploids and their diploid parents. The PCR primers of successful validation are listed in Table S7.

Statistics

All Statistical tests in this paper were performed using basic packages in R (Version 3.3.1, https://www.r-project.org/).

Results

Using the synthetic rice segmental allotetraploids generated by inter-subspecies hybridization and colchicine-mediated whole-genome duplication (WGD) between rice subspecies, *japonica* (cv. Nipponbare) and *indica* (cv. 9311) (Xu et al. 2014; Sun et al. 2017), we characterized, compared, and correlated the inter-subgenomic homoeologous exchanges, changes in gene expression, and alternative splicing profiles in individual progeny of the rice segmental allotetraploids.

Extensive genomic homogenization via homoeologous exchanges in the rice segmental allotetraploids

Inspired by the fact that vast genomic variation could be induced by allopolyploidization in plants (Doyle et al. 2008), we initially characterized the genomic DNA composition of rice segmental allotetraploid individuals (designated Tetra 1 through Tetra 4) relative to their diploid parents. The re-sequencing results showed that sequencing depth was uniformly distributed among all reference chromosomes in each tetraploid individual (Figure S1), indicating all four rice segmental allotetraploid individuals were euploids, which eliminated the potential confounding effects of aneuploidy on subsequent analyses (Wu et al. 2018).

We characterized the inter-subgenome homoeologous exchanges (abbreviated as HEs) via comparing the genomic compositions in each of the four tetraploid individuals relative to their diploid parental genomes ("Materials and methods"). Accordingly, as illustrated and summarized in Fig. 1a, b, approximately 70–82% of genomic regions in the S_{10} tetraploid individuals were homogenized, including 29–34% Nipponbare-homogenized regions (abbreviated as NPB-HG regions) and 40–48% 9311-homogenized regions (abbreviated as 9311-HG regions; Fig. 1a, b); the rest 18–30% genomic regions were still in a heterozygous state with biparental contributions (Fig. 1a, b). Notably, the average proportion of NPB-HG regions was significantly lower than that of 9311 across all four tetraploids (p value < 0.01, by Paired t test, Fig. 1b).

Further analysis of genomic composition of the tetraploids revealed the following interesting features. First, different chromosomes harbored different proportions of homogenized and heterozygous regions (Fig. 1a). For example, Chr 10 maintained almost complete biparental heterozygosity over its entire length across all four tetraploids,
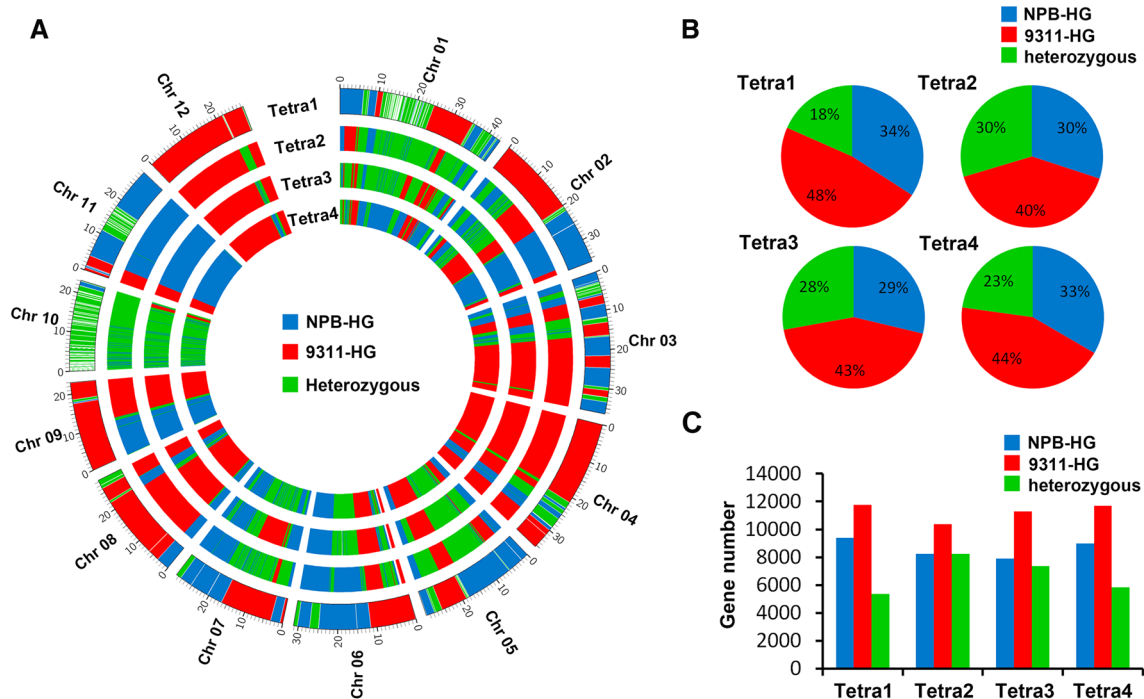
Fig. 1 Genomic constitution of rice tetraploid individuals. **a** Circus plots depicting the genomic constitution of four tetraploid individuals along each of the 12 rice chromosomes. Blue, red and green represent Nipponbare-homogenized (NPB-HG), 9311-homogenized (9311-HG) and heterozygous regions, respectively. From the outer circle to the inner circle represent Tetra 1 to Tetra 4, respectively. **b** Pie charts showing genotypes of each tetraploid individual. **c** Distribution of 26,529 high-confidence genes in different types of genomic regions for downstream analysis. Representation of different colors in both (**b**) and (**c**) were same as those in (**a**) (color figure online)

while Chr 8, Chr 11 and Chr 12 and part of Chr 4 showed near complete homozygosity to either NPB (NPB-HG) or 9311 (9311-HG), across all four tetraploids. Second, parental homogeneities were largely conserved among the Tetra 2 through Tetra 4 tetraploids in terms of their distribution patterns; however, Tetra 1 was always the outlier, in which almost each chromosome frequently displayed inconsistent homogenization patterns from the other three tetraploids. For example, Chr 09 in Tetra 1 mainly harbored homogenized regions from 9311, whereas such regions in Tetra 2 through Tetra 4 showed 9311-HG and NPB-HG over approximately half of the chromosome.

To investigate whether the extensive HEs could affect the homoeologous gene composition in the tetraploids, we categorized the genes into groups in terms of their genomic locations within homogenized and heterozygous regions. Regarding the variable gene models for each gene, we focused on 26,529 non-transposon-related gene models with significant sequence conservation in their genic exons between NPB and 9311 in categorization and following analyses ("Materials and methods"). Notably, the relative genic proportions of 9311-HG, NPB-HG and heterozygous genes were similar to the corresponding genomic proportions of parental homogeneity and heterozygosity in the respective tetraploid individuals (Fig. 1c).

Majority of the homogenized genes in the tetraploids showed similar expression as their diploid parental counterparts but some manifested novel expression

In addition to the extensive genomic homogenization induced by HEs in the tetraploids, we next analyzed the expression of genes within the homogenized genomic regions relative to the expression of their counterparts in the corresponding parent (NPB-HG regions vs. counterpart homologs in NPB and 9311-HG regions vs. counterpart homologs in 9311). Specifically, we employed RNA-sequencing to characterize the transcriptome profiles in the embryos tissue of the four tetraploids as mentioned in the previous sections. Within each tetraploid individual, expression of the genes residing within the NPB-HG and 9311-HG regions was quantified and compared with those of their homologs in parental NPB and 9311, respectively. Accordingly, the genes with significantly altered expression levels relative to their corresponding diploid parental homologs in each tetraploid individual (revealed by tetraploid vs. corresponding parent) were defined as differentially expressed genes (DEGs; detailed in "Materials and methods"), whereas the remaining genes were defined as non-DEGs (Fig. 2a). For non-DEGs, regardless of their expression levels being
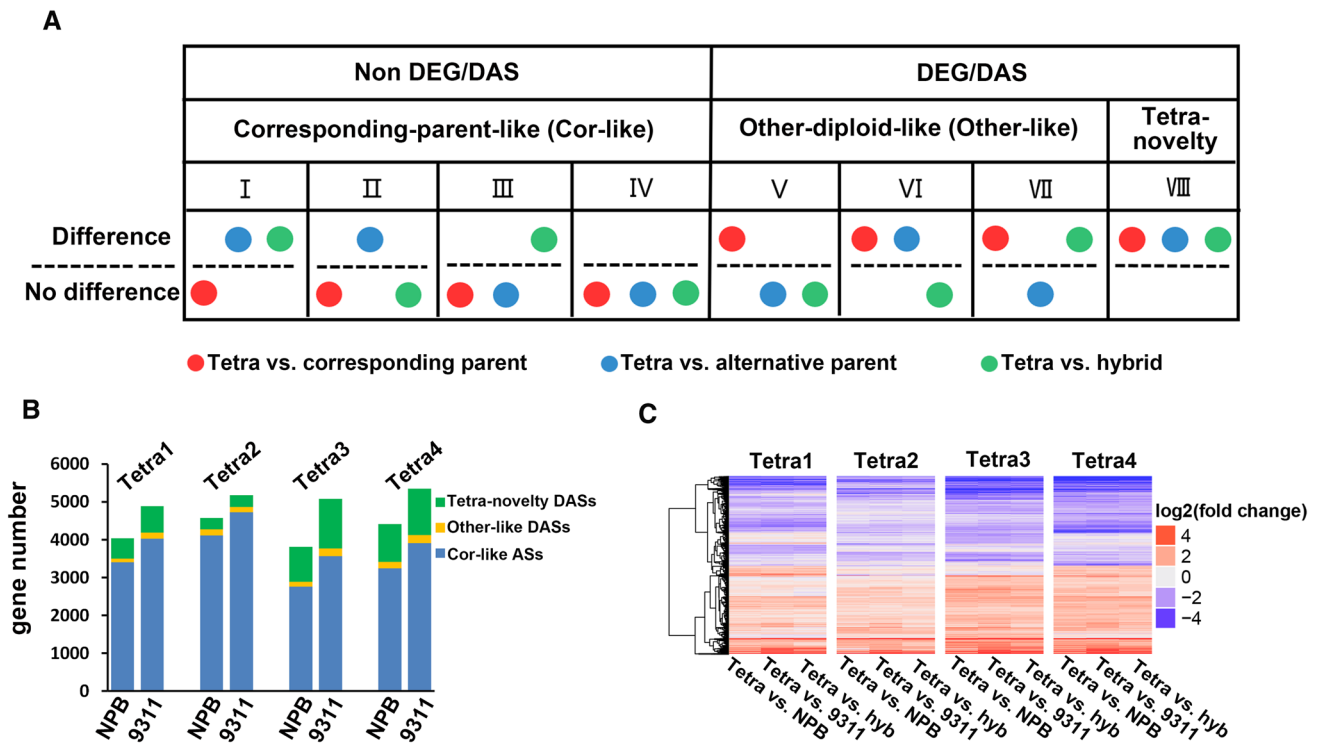
Fig. 2 Expression profile of genes located in homogenized regions (NIP-HG and 9311-HG regions) in rice segmental allotetraploids. **a** The eight possible genes expression/alternative splicing (AS) patterns in homogenized regions of segmental allotetraploids relative to their corresponding diploid parent were divided into three groups: corresponding parent like genes/AS events (I–IV, denoted as cor-like genes/ASs), other-diploid-like DEGs/DASs (V–VII, denoted as other-like DEGs/DASs) and tetraploid-novelty DEGs/DASs (VIII, denoted as tetra-novelty DEGs/DASs). Circles with different colors represent different comparisons between tetraploid and diploid. The dashed line in each state represents the cutoff of comparisons: Circles above the dashed lines means significant differences in a certain comparison, and circles below the dashed lines mean no difference in a certain comparison. **b** Proportion of genes in homogenized regions (NPB-HG and 9311-HG genes) belonged to cor-like, other-like and tetra-novelty groups in tetraploid individuals. **c** Hierarchical clustering of merged tetra-novelty genes in all four tetraploids based on expression changes between tetraploids and diploids, red and purple represent up- and down-regulated expression changes, respectively (color figure online)

similar to or different from the alternative parent and/or F1 hybrid, they were generally characterized as the genes with "corresponding-parent-like" genes (abbreviated as cor-like genes; categories I-IV in Fig. 2a); for DEG genes, two distinct groups were partitioned: (i) DEGs showed similar expression level with either alternative parent or F1 hybrid were defined as "other-diploid-like" DEGs (abbreviated as others-like DEGs; categories V-VII in Fig. 2a; Figure S2) and (ii) the remaining DEGs were defined as "tetraploid-novelty" DEGs (abbreviated as tetra-novelty DEGs), which were differently expressed when compared with both the alternative parent and the F1 hybrid (Fig. 2a, category VIII and Figure S2).

We found that only 916-2566 (9.4–28.9%) of the genes located in homogenized regions became DEGs in the tetraploids compared with their parental counterparts (Table S1), implicating that most homogenized genes maintained the expression levels similar to their respective diploid parents despite the change in ploidy, suggesting they were largely under *cis*-regulation. When the DEGs were divided based

on their parental origins, we found that different tetraploid individuals represented inconsistent parental-biased DEGs (Table S2, Exact binomial test, p value < 0.05), which means no parental bias for DEGs was found. Intriguingly, the majority of the DEGs (66.9–86.8%) belonged to the tetra-novelty group (Fig. 2b, Figure S1 and Table S1). To explore the expression consistency of "tetra-novelty DEGs" across tetraploids, we took the combined set of all "tetra-novelty DEGs" from each tetraploid individual and calculated the fold changes in expression relative to their diploid parents and F1 hybrids. As Fig. 2c illustrates, the total tetra-novelty DEGs were mostly clustered into two distinct groups involving concordant up- and down-regulated gene sets, whereas only 0.5% tetra-novelty DEGs (17 out of 3772) showed inconsistent expression patterns among the tetraploids, i.e., up-regulated in some individuals and down-regulated in others and vice versa. This implicates that the expression patterns of tetra-novelty DEGs were controlled by conserved *trans*-acting factors and may have occurred non-randomly across tetraploid individuals.

Furthermore, to explore the potential functional relevance of the tetra-novelty genes, we performed gene ontology (GO) analysis for all tetra-novelty genes that showed such expression in at least one of the tetraploid individuals. We found these genes were enriched in pathways related to abiotic stresses (for both up-regulated and down-regulated tetra-novelty genes) and metabolic processes (mainly consists of down-regulated tetra-novelty genes) (Figure S3). Notably, the biological processes such as mRNA splicing (19 genes), RNA secondary structure unwinding (18 genes) and mRNA catabolic process (5 genes) were also enriched by the up-regulated tetra-novelty genes, which indicates a possible connection between up-regulated tetra-novelty genes and post-transcriptional alternative splicing processes (Kornblihtt et al. 2013; Smith and Baker 2015).

Enhanced alternative splicing was associated with the homogenized genes in the tetraploids relative to their diploid parental counterparts

Given that polyploidy can result in changes in alternative splicing (AS) events (Zhou et al. 2011), we analyzed AS in the rice tetraploids. Based on the same RNA-sequencing data used for expression analysis, we surveyed AS patterns of all homogenized genes in the rice tetraploids and their counterparts in the corresponding diploid parent. Based on the pipeline reported by Wang (Wang et al. 2016) with several modifications, 9,467 curated AS events derived from 5,373 genes were identified in all assayed samples, which can be classified into four categories: intron retention (IR), exon skipping (ES), alternative donor (AltD), and alternative acceptor (AltA) (Fig. 3a and Table S3; detailed in "Materials
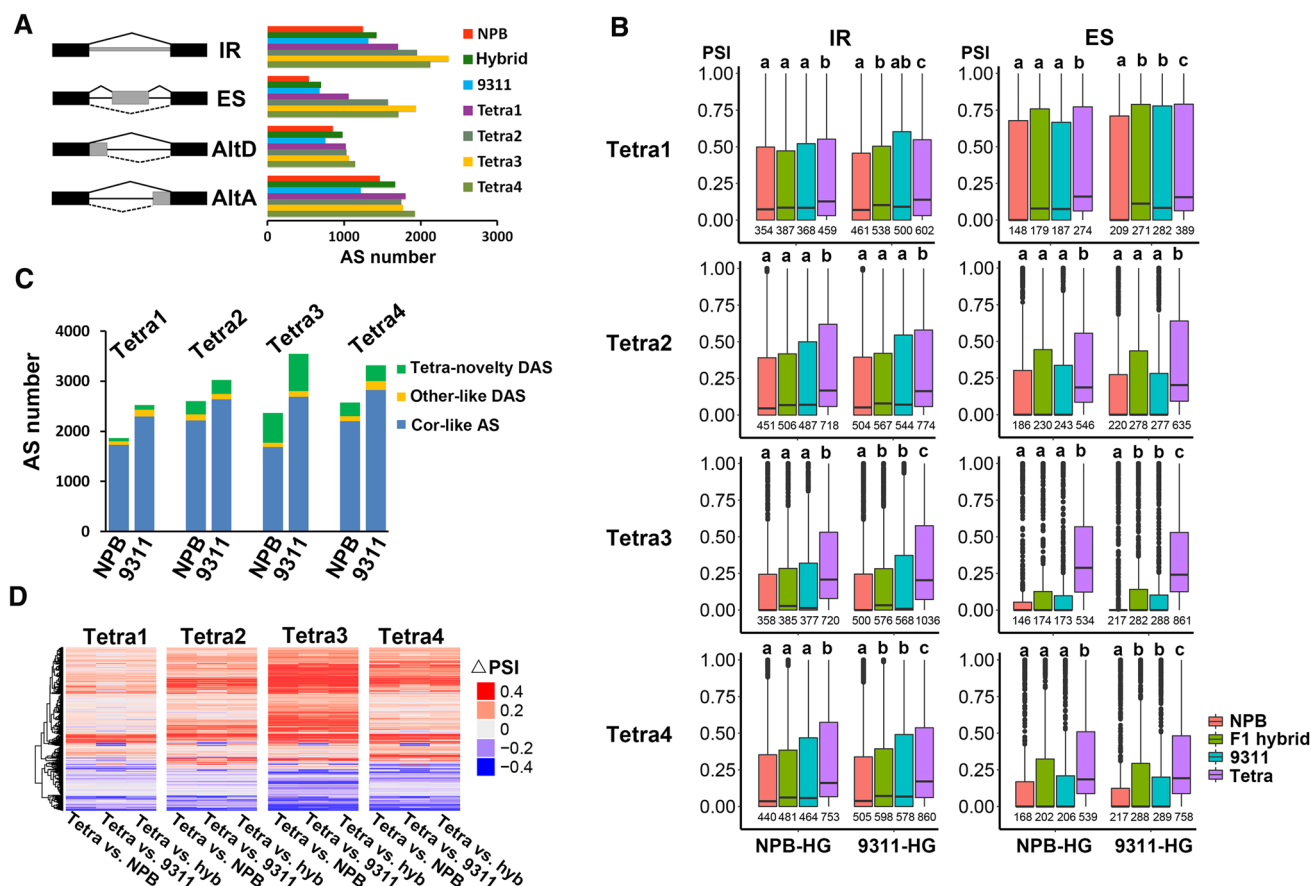

Fig. 3 Changes in AS events in homogenized genes relative to corresponding diploid parents. **a** Different types of AS events and their occurrence in diploids and tetraploids. **b** Comparisons of PSI values of intron retention (IR) and exon skipping (ES) events among homogenized regions of tetraploids (purple) and their counterparts in corresponding parents (NPB in red, 9311 in green and the F1 hybrid in blue, respectively). For each small panel, the left part represents NPB-HG regions, whereas the right part represents 9311-HG regions. Letters above each box represent statistically different PSI distributions in each comparison. The numbers under each box represent the relevant number of AS events. **c** Proportion of AS events in homogenized regions (NPB-HG and 9311-HG AS events) belonged to cor-like (blue), other-like (yellow) and tetra-novelty (green) groups. **d** Hierarchical clustering of all merged tetra-novelty AS events in different tetraploids based on changes in AS expression level between tetraploids and diploids, red and blue represents AS with up- and down-regulated changes, respectively (color figure online)

and methods"). Similar to previous studies in other plant species (Li et al. 2014; Liu et al. 2017), IR was the most common AS events (36%), followed by ES (26%), AltA (24%) and AltD (14%) (Tables S3). Authenticity of these identified IRs (not due to sequencing of immature pre-mRNA (Wang et al. 2018a) was supported by the significant higher ratio of complete ORFs in IR-related novel transcripts (3089 in 3871) than that in all background novel transcripts harboring other AS events (8661 in 11557) (p value $= 5.872e^{-10}$ by Fisher's exact test, "Materials and methods"). Intriguingly, more AS events were found in the homogenized genes in the tetraploids than their counterparts in the diploid parents and F1 hybrid (Fig. 3a). To explore whether the strength of splicing signal was also elevated after WGD and HEs, the Percentage of Splicing Index (PSI, the ratio between reads including or excluding exons, indicating the efficiency of splicing a given exon into all transcripts of a gene) (Schafer et al. 2015) of AS events occurred in the homogenized genes in each tetraploid individual and diploids are tabulated in Fig. 3b and Figure S4. Notably, for all types of AS events that occurred in homogenized genes, the tetraploids showed larger PSI values (p values < 0.01 by Kruskal–Wallis post hoc test) and increased AS (p values < 0.05 by Permutation test) compared to their corresponding parent (Fig. 3b and Figure S4). However, taking AS in the alternative parent and F1 hybrid into account, AltD and AltA mostly showed higher PSI than those in the alternative parent but similar with those in the F1 hybrid, whereas IR and ES remained the highest in the tetraploids. Furthermore, both the AS numbers and PSI values of splicing-related genes, including Serine/Arginine-rich (SR), heterogeneous nuclear RNP (hnRNP) genes and other splicing-related factors were significantly higher in tetraploids than those in diploids (Figure S5), probably due to an enriched repository of *trans*-acting splicing factors in rice tetraploids after WGD and HEs.

To characterize the altered AS profiles of the homogenized genes in tetraploids, we adopted similar categorizing methods utilized in gene expression analysis. AS events that showed significant differences (detailed in "Materials and methods") in homogenized genes relative to their corresponding diploid parental counterparts were defined as the Differential AS events (DAS). Likewise, corresponding-parent-like AS events (abbreviated as cor-like AS), other-diploid-like DAS events (abbreviated as other-like DAS) and tetraploid-novelty DAS events (abbreviated as tetra-novelty DAS) were defined similarly as the aforementioned gene expression (Fig. 2a). In tetraploids, most homogenized genes exhibited similar AS patterns to their corresponding diploid counterparts (71.2–92.4% ASs in NPB-HG regions and 75.6–91.1% ASs in 9311-HG regions). However, there were 368-1,543 DAS (7.6–28.8% DAS in NPB-HG regions and 8.9–24.3% DAS in 9311-HG regions, respectively) that occurred in the tetraploids only (Fig. 3c, Figure S6 and

Table S4). Analogous to DEGs, no bias was detected for DAS between NPB-HG and 9311-HG regions (Table S5). Notably, tetra-novelty DAS also had the highest percentage class for all DAS (Fig. 3c) and which clustered into two distinct groups (up- and down-regulated DAS events) with more up-regulated DAS across all tetraploids (Fig. 3d).

We further dissected the tetra-novelty DAS into two subsets including up-regulated and down-regulated AS by comparing tetraploids with the corresponding diploid parents. For IR, ES, and AltA, tetra-novelty DAS harbored more up-regulated DAS than down-regulated DAS among all tetraploids, whereas those belonging to AltD showed the opposite trend (Figure S7 and Table S6). With dramatic changes in AS, IR and ES showed two interesting sequence characteristics: (i) introns involved up-regulated IR events were shorter and had higher GC contents than all IRs and constitutive introns; and (ii) exons involved up-regulated ES events were shorter than all ESs and constitutive exons, whereas the GC contents were higher than constitutive exons but identical with all ESs (Figure S8). Furthermore, we performed AS-specific RT-PCR by using six randomly chosen up-regulated tetra-novelty DAS transcripts for IR and ES classes. We found that in two IR and four ES genes, the AS events (isoform 2) either occurred exclusively or more abundantly in the tetraploids compared with the diploid parents and the F1 hybrid, thus validating our RNA-seq analysis results (Figure S9).

Expression change and alternative splicing of the homogenized genes in the tetraploids are modulated independently

Previous studies in diploid plant species have indicated that there were no obvious associations between changes in AS signaling and transcriptional activity in response to biotic and abiotic stresses (Chang et al. 2012; Ling et al. 2015). To test whether this also applies to the rice tetraploids, we examined the relationship between DEGs and DAS-related genes (DSGs). Initially, all interested DSGs located on homogenized regions (5,373 homogenized genes) in each tetraploid individual were assigned into specific groups based on their gene expression and AS patterns, including the DEG-DSG group (genes both exhibiting differential expression patterns [DE] and meanwhile displaying DAS), DEG-only (genes only exhibiting DE), DSG-only (genes only showing DAS), and neither group (genes neither exhibiting DE nor showing DAS). Genes belonging to each group are tabulated in Fig. 4a. There were a limited number of genes (27-238) categorized into the DEG-DSG group in the tetraploids (Fig. 4). Additionally, it was reported that the transcript abundance of pre-mRNA may affect splicing processes (Kornblihtt et al. 2013; Liu et al. 2017). To test this, we compared the proportions of DSGs in both DEGs and non-DEGs. We found that DSGs were mainly
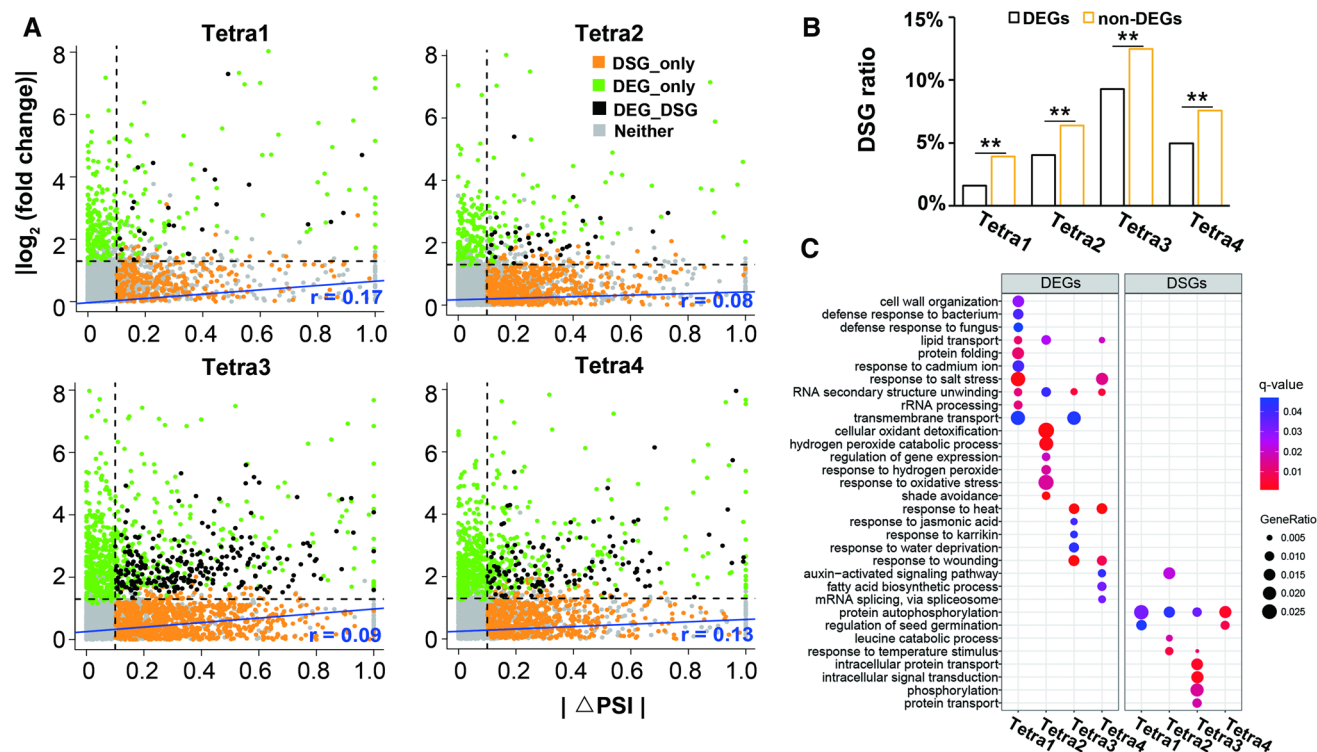
Fig. 4 Relationship between DEGs and DSGs (DAS-related genes) in homogenized regions of the four tetraploids. **a** Scatter plots depict the gene categories of DEG-only (orange), DSG-only (green), both DEG and DSG (black) and neither DEG nor DSG (gray) in each tetraploid. The x-axis represents $\log_2$ transformed fold change of gene expression, and the y-axis represents ΔPSI between tetraploids versus corresponding diploids, respectively (for genes possessing multiple DAS transcripts, the maximum of ΔPSI is used as the agent; for genes without any DAS, the minimum of ΔPSI is used as the agent). Blue line represents the liner regression between the delta of DEG and DSG with respective correlation coefficient r values denoted. **b** Comparison of DSG ratio between DEGs and non-DEGs. Asterisks represent significant differences by Fisher's exact test (p value <0.01). **c** GO analysis of both DEGs and DSGs in tetraploids. The color scale represents FDR-corrected p values derived from hypergeometric distribution tests, and the size scale of dots symbolizes the ratio of enriched genes in all DEGs or DSGs (color figure online)

enriched in non-DEGs rather than DEGs (Fisher's exact test, p value <0.01; Fig. 4b), suggesting there is no association between transcript abundance and AS.

The functional relationship between DEGs and DSGs was also explored based on GO enrichment analysis. We found that DEGs were mainly associated with biological processes involved in biotic and abiotic stresses, whereas DSGs were enriched in more specific GO terms such as protein autophosphorylation and regulation of seed germination (Fig. 4c). Collectively, the limited number of DEG-DSG group along with very little commonality in enriched GO terms between DEGs and DSGs strongly suggest independent modulation of gene expression and AS of homogenized genes after WGD and HEs in rice segmental allopolyploids.

Discussion

Allopolyploidy or segmental allopolyploidy, involving the merger and doubling of diverged genomes, can result in various types of extensive genomic variation (Szadkowski et al. 2011; Chalhoub et al. 2014; Li et al. 2015). Among these, homoeologous exchanges (HEs), featured as mutual exchanges of genomic sequences via meiotic homoeologous recombination, have been well-characterized in several different allopolyploid plant systems (Kovarik et al. 2012; Henry et al. 2014; Soltis et al. 2016; Lloyd et al. 2017). In allopolyploid plants, most research on the potential effects of HEs on genomic composition mainly focused

on exploring the relative dosages of subgenomic and/or genic sequences (Chalhoub et al. 2014; Li et al. 2015). Other studies focused on alternations in gene expression caused by allopolyploidy and mainly explored the non-additive and/or biased expression of homoeologous genes at the transcriptional level (Buggs et al. 2011; Grover et al. 2012; Xu et al. 2014) and alternative splicing (AS) changes at the post-transcriptional level (Zhou et al. 2011; Liu et al. 2017) resulting from WGD. However, the effects of HEs on gene expression and AS are still poorly understood. In this study, we used selfed progeny of a synthetic segmental allotetraploid rice system to explore the association between HEs, gene expression and AS with a focus on homogenized genic regions.

Frequently, HEs within the homologous genomic regions of allopolyploids were reported in allotetraploid species of *Brassica* and *Tragopogon* (Chester et al. 2012; Chalhoub et al. 2014; He et al. 2016). In fact, when parental genomic divergence is low, there is an increase in HEs in similar subgenomic regions in *B. napus* allopolyploids generated by meiotic homoeologous recombination (Gaeta and Pires 2010). In our S_{10} rice tetraploids, between 70%-82% of the genome was converted to homogeneity from heterozygous regions through the occurrence of extensive HEs, while on average *ca*. 50% were homozygous in S5 (Li et al. 2019), suggesting rapid generational increment of genome-wide homozygosity. The strikingly vast and widespread HEs in these rice segmental allotetraploids are probably due to the occurrence of more crossing-overs between homoeologous chromosomes and the accumulation of HEs in advanced generations. These results are consistent with high genomic similarity between the two parental rice subspecies, *japonica* and *indica* (Tsunematsu et al. 1996; Yang et al. 2012). The remnant of heterozygous regions in the 10th-generation tetraploids implicates that HE is a still ongoing process in our rice segmental tetraploid system. However, it is difficult to anticipate how many more generations are required to render all genomic regions homogenized or fixed with certain homoeologous subgenomes.

The regulation of gene expression involves both *cis* regulators (i.e., the promoter, enhancer, insulator, etc.) and *trans* factors (i.e., the transcriptional factor, etc.) in eukaryotic cells (Lienert et al. 2011; Shi et al. 2012; Wittkopp and Kalay 2012). In our case of segmental allotetraploid rice, especially for homogenized genes, the *cis*-regulation (parental legacy) was determined to be dominant since the majority of genes were categorized into the corresponding-diploid-like group. Notably, there were instances where despite genic regions retaining the same homogenized genomic constitution as their parent, a subset of genes was not expressed as in the corresponding diploid parent. One explanation may be that allopolyploidization could aggregate diverged *trans*-acting regulatory elements and eventually lead to the novel

patterns of gene activation or repression in novel regulatory environments as has been documented in other systems (Grover et al. 2012; Shi et al. 2012; Combes et al. 2013). The vast majority of differentially expressed genes occurred in tetraploids, which implicates that differential expression patterns were established over the course of genome doubling during allopolyploidization and subsequent HEs rather than in the genome hybridization process. In addition, conserved expression remodeling of tetra-novelty DEGs across four tetraploid individuals also supports the assertion that novel *trans*-regulators were conservatively controlled by certain molecular mechanisms rather than mediated by stochastic effects. Taken together, we found that homogenized gene homoeologous within rice segmental allotetraploids mostly retained *cis*-regulatory elements to maintain the original parental expression pattern; however, the whole genome duplication and subsequent genomic rearrangements were capable of changing the stability of the transcriptome via inducing variation in *trans*-regulatory factors. In addition, the extensive regional repatterning of DNA methylation in earlier generations of these tetraploids (Li et al. 2019) should also be an important contributing factor to changes of gene expression.

Previous studies showed that AS events were obviously induced in response to abiotic/biotic stresses in plants (Feng et al. 2014; Vitulo et al. 2014; Ling et al. 2015; Liu et al. 2017). Polyploidization could also promote the change of AS events and signals as the responses to the changed nuclear environment involving whole-genome duplication (Zhou et al. 2011; Saminathan et al. 2015). Our results provided solid evidence that segmental allopolyploidy and subsequent HEs could not only inherit the AS events generated by hybridization (AltD and AltA events) but also induce more novel AS events and enhanced AS signals within the homogenized regions of tetraploids (novel IR and ES events). Interestingly, we found that exon skipping was the second most common type of AS in our rice polyploids, which was consistent with previous studies in rice and wheat (Li et al. 2014; Liu et al. 2017). Accordingly, potential variable genic features in respective species, such as transcriptional level and exon–intron architecture, may affect the composite profile of the AS events (Keren et al. 2010). The increase in AS events and enhanced AS signals were probably due to the altered AS of the pre-mRNA of SRs, hnRNPs and other splicing-related genes, whose activity and abundance was previously proposed to determine the AS profiles of their target genes (Syed et al. 2012; Reddy et al. 2013). As expected, splicing-related genes indeed showed significant AS changes in the tetraploids relative to the diploid parents, which may play important roles for transcript diversification in our rice tetraploids.

Furthermore, most AS isoforms of homogenized genes in tetraploids were corresponded to the parental AS

isoforms, which implies that some potential *cis*-acting elements such as splicing enhancers (SEs), splicing silencers (SSs) (Kornblihtt et al. 2013) and even the intrinsic genic structural features (i.e., the length and GC content of genic introns and exons) were able to maintain the overall parental AS profiles. Moreover, certain *cis*-acting intrinsic genic structural features, such as the length and GC content of genic introns and exons, could also contribute to the AS remodeling in rice allotetraploids. In addition, *trans*-acting factors also participated in the remodeling of post-transcriptional AS profiles even within those homogenized genes since there was still a nonnegligible proportion of differential alternative splicing events that mainly arise in the tetraploids.

Moreover, since a limited number of differentially expressed genes and differential alternatively spliced genes share commonality in enriched GO terms in our rice allotetraploids, it implicates that transcriptional regulation and AS are modulated independently in response to whole-genome doubling and further homogenization via HEs. This result is consistent with the conclusions from previous studies that focused on differences in AS in response to heat stress, high salinity and insect herbivory (Chang et al. 2012; Feng et al. 2014; Ling et al. 2015). In this aspect, it is interesting to note that both gene expression and AS can be impacted by changes of DNA methylation (Wang et al. 2016).

Collectively, HE represents a major mechanism to induce phenotypic innovation in nascent allopolyploid plants via generating novel genomic composition, transcriptional variation in gene expression, and alternative spliced transcripts, which may provide new insights into allopolyploid crop breeding.

Author contribution statement Z.B.Z., B.L., Y.W. and L.G. designed the research. Z.B.Z., T.S.F., Y.W. and L.G. performed the research. Z.B.Z., T.S.F., Z.J.L., X.T.W. and B.X.D. analyzed data. H.W.X. and G.L. did experimental validation. Z.B.Z., B.L., Y.W. and L.G. wrote the manuscript. Y.Z.D, X.Y.L. and K.A.S. modified the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc 1:2320–2325

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:1–12

Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169

Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Curr Biol 21:551–556

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345:950–953

Chang CY, Lin WD, Tu SL (2012) Genome-wide analysis of heat-sensitive alternative splicing in *Physcomitrella patens*. Plant Physiol 165:826–840

Chester M, Gallagher JP, Symonds VV, Silva AVCD, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc Natl Acad Sci USA 109:1176–1181

Combes MC, Dereeper A, Severac D, Bertrand B, 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

Doyle JJ, Doyle JL, Rauscher JT, Ahd B (2003) Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (Glycine subgenus Glycine). New Phytol 161:121–132

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–461

Feng D, Peng C, Wang Z, Zhang S, Ali S, Xiong L (2014) Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. BMC Genom 15:431

Filichkin SA, Cumbie JS, Dharmawardhana P, Jaiswal P, Chang JH, Palusa SG, Reddy AS, Megraw M, Mockler TC (2015) Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in *Arabidopsis*. Mol Plant 8:207–227

Flagel LE, Wendel JF, Udall JA (2012) Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genom 13:302

Foissac S, Sammeth M (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res 35:297–299

Gaeta RT, Pires JC (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186:18–28

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF (2012) Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol 196:966–971

Guo X, Han F (2014) Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. Plant Cell 26:4311–4327

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512

He Z, Wang L, Harper AL, Havlickova L, Pradhan AK, Parkin IAP, Bancroft I (2016) Extensive homoeologous genome exchanges

in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnol J 15:594–604

Henry IM, Tyagi A, Comai L (2014) The BOY NAMED SUE quantitative trait locus confers increased meiotic stability to an adapted natural allopolyploid of *Arabidopsis*. Plant Cell 26:181–194

Hurgobin B, Golicz AA, Bayer PE, Chan CK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin I (2017) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol J 16:1265–1274

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100

Keren H, Levmaor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11:345–355

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol 14:153–165

Kovarik A, Renny-Byfield S, Grandbastien MA, Leitch A (2012) Evolutionary implications of genome and karyotype Restructuring in *Nicotiana tabacum* L. In: Soltis P, Soltis D (eds) Polyploidy and genome evolution. Springer, Berlin, Heidelberg

Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K (2010) Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. Genetics 186:801–812

Lashermes P, Combes MC, Hueber Y, Severac D, Dereeper A (2014) Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. Plant J 78:674–685

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map (SAM) format and SAMtools. Transpl Proc 19:1653–1654

Li Q, Xiao G, Zhu YX (2014) Single-nucleotide resolution mapping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. Mol Plant 7:829–840

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol 33:524–530

Li N, Xu CM, Zhang A, Lv RL, Meng XC, Lin XY, Gong L, Wendel JF, Liu B (2019) DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. New Phytologist. https://doi.org/10.1111/nph.15820

Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D (2011) Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 43:1091–1097

Ling Z, Zhou W, Baldwin IT, Xu S (2015) Insect herbivory elicits genome-wide alternative splicing responses in *Nicotiana attenuata*. Plant J Cell Mol Biol 84:228–243

Liu L, Stein A, Wittkop B, Sarvari P, Li J, Yan X, Dreyer F, Frauen M, Friedt W, Snowdon RJ (2012) A knockout mutation in the lignin biosynthesis gene CCR33 explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. Theor Appl Genet 124:1573–1586

Liu Z, Qin J, Tian X, Xu S, Wang Y, Li H, Wang X, Peng H, Yao Y, Hu Z (2017) Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum asetivum* L.). Plant Biotechnol J 16:714–726

Lloyd A, Blary A, Charif D, Charpentier C, Tran J, Balzergue S, Delannoy E, Rigaill G, Jenczewski E (2017) Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. New Phytologist 217:367–377

Madlung A, Wendel JF (2013) Genetic and epigenetic aspects of polyploid evolution in plants. Cytogenet Genome Res 140:270–285

Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C (2014) Functional consequences of splicing of the antisense transcript *COOLAIR* on *FLC* transcription. Mol Cell 54:156–165

Pertea M, Kim D, Pertea G, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. Nat Protoc 11:1650–1667

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842

Reddy AS, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. Plant Cell 25:3657–3683

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Bot 101:1711–1725

Rosloski SM, Singh A, Jali SS, Balasubramanian S, Weigel D, Grbic V (2013) Functional analysis of splice variant expression of *MADS AFFECTING FLOWERING 2* of *Arabidopsis* thaliana. Plant Mol Biol 81:57–69

Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol 186:123–134

Saminathan T, Nimmakayala P, Manohar S, Malkaram S, Almeida A, Cantrell R, Tomason Y, Abburi L, Rahman MA, Vajja VG (2015) Differential gene expression and alternative splicing between diploid and tetraploid watermelon. J Exp Bot 66:1369–1385

Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N (2015) Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). Curr Protoc Hum Genet 24:11–11

Seo PJ, Park MJ, Lim MH, Kim SG, Lee M, Baldwin IT, Park CM (2012) A self-regulatory circuit of *CIRCADIAN CLOCK-ASSOCIATED1* underlies the circadian clock regulation of temperature responses in *Arabidopsis*. Plant Cell 24:2427–2442

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

Smith JE, Baker KE (2015) Nonsense-mediated RNA decay: a switch and dial for regulating gene expression. BioEssays 37:612–623

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. Curr Opin Genet Dev 35:119–125

Soltis DE, Visger CJ, Marchant DB, Soltis PS (2016) Polyploidy: pitfalls and paths to a paradigm. Am J Bot 103:1146–1166

Song Q, Chen ZJ (2015) Epigenetic and developmental regulation in plant polyploids. Curr Opin Plant Biol 24:101–109

Stebbins GL (1947) Types of polyploids; their classification and significance. Adv Genet 1:403–429

Sturgill D, Malone JH, Sun X, Smith HE, Rabinow L, Samson ML, Oliver B (2013) Design of RNA splicing analysis null models for post hoc filtering of *Drosophila* head RNA-seq data with the splicing analysis kit (Spanki). BMC Bioinform 14:1–18

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960

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

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants–coming of age. Trends Plant Sci 17:616–623

Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre AM (2011) Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. New Phytol 191:884–894

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2014) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578

Tsunematsu H, Yoshimura A, Harushima Y, Nagamura Y, Kurata N, Yano M, Sasaki T, Iwata N (1996) RFLP framework map using recombinant inbred lines in rice. Breed Sci 46:279–284

Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. Nat Rev Genet 18:411–424

Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, Zimbello R, Corso M, Vannozzi A, Bonghi C (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol 14(2):99

Wang X, Hu L, Wang X, Li N, Xu C, Gong L, Liu B (2016) DNA methylation affects gene alternative splicing in plants: an example from rice. Mol Plant 9:305–307

Wang J, Symul L, Yeung J, Gobet C, Sobel J, Lück S, Westermark PO, Molina N, Naef F (2018a) Circadian clock-dependent and -independent posttranscriptional regulation underlies temporal mRNA accumulation in mouse liver. Proc Natl Acad Sci USA 115:E1916–E1925

Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, Pei L, Wang F, Hu J, Tu L (2018b) A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. New Phytol 217:163–178

Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13:59–69

Wu Y, Sun Y, Sun S, Li G, Wang J, Wang B, Lin X, Huang M, Gong Z, Sanguinet KA, Zhang Z, Liu B (2018) Aneuploidization under segmental allotetraploidy in rice and its phenotypic manifestation. Theor Appl Genet 131(6):1273–1285

Xu F, Xu S, Wiermer M, Zhang Y, Li X (2012) The cyclin L homolog MOS12 and the MOS4-associated complex are required for the proper splicing of plant resistance genes. Plant J Cell Mol Biol 70:916–928

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

Yang CC, Kawahara Y, Mizuno H, Wu J, Matsumoto T, Itoh T (2012) Independent domestication of Asian rice followed by gene flow from japonica to indica. Mol Biol Evol 29:1471

Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R Package for comparing biological themes among gene clusters. Omics J Integr Biol 16:284–287

Zhang H, Bian Y, Gou X, Zhu B, Xu C, Qi B, Li N, Rustgi S, Zhou H, Han F (2013) Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proc Natl Acad Sci USA 110:3447–3452

Zhao J, Udall JA, Quijada PA, Grau CR, Meng J, Osborn TC (2006) Quantitative trait loci for resistance to *Sclerotinia sclerotiorum* and its association with a homeologous non-reciprocal transposition in *Brassica napus* L. Theor Appl Genet 112:509–516

Zhou R, Moshgabadi N, Adams KL (2011) Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc Natl Acad Sci USA 108:16122–16127

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Zhibin Zhang[1] · Tiansi Fu[1] · Zhijian Liu[1] · Xutong Wang[2] · Hongwei Xun[1] · Guo Li[1] · Baoxu Ding[1] · Yuzhu Dong[1] · Xiuyun Lin[3] · Karen A. Sanguinet[4] · Bao Liu[1] · Ying Wu[1]⊙ · Lei Gong[1]

[1] Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun 130024, China

[2] Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

[3] Jilin Academy of Agricultural Sciences (JAAS), Changchun 136100, China

[4] Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA