Asymmetric Divergence in Transmitted SNPs of DNA Replication/Transcription and Their Impact on Gene Expression in Polyploid *Brassica napus*

Minqiang Tang[1], Juanling Li[1], Xu Hu[1], Lu Sun[1], MMU Helal[2], Jianguo Chen[3]* and Yuanyuan Zhang[2]*

[1]*Key Laboratory of Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants (Ministry of Education), College of Forestry, Hainan University, Haikou, China,* [2]*The Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture, Oil Crops Research Institute of Chinese, Academy of Agricultural Sciences, Wuhan, China,* [3]*School of Life Sciences, Hubei University, Wuhan, China*

The marked increase in plant genomic data has provided valuable resources for investigating the dynamic evolution of duplicate genes in polyploidy. *Brassica napus* is an ideal model species for investigating polyploid genome evolution. The present study comprehensively analyzed DNA and RNA variation of two representative *B. napus* inbredlines, Zhongshuang11 and Zhongyou821, and we investigated gene expression levels of A_n and C_n subgenomes in multiple tissues of the two lines. The distribution of transmitted single nucleotide polymorphisms (SNPs) was significantly different in two subgenomes of *B. napus*. Gene expression levels were significantly negatively correlated with number of variations in replication and transcription of the corresponding genes, but were positively correlated with the ratios of transmitted SNPs from DNA to RNA. We found a higher density of SNP variation in A_n than that in C_n during DNA replication and more SNPs were transmitted to RNA during transcription, which may contribute to A_n expression dominance. These activities resulted in asymmetrical gene expression in polyploid *B. napus.* The SNPs transmitted from DNA to RNA could be an important complement feature in comparative genomics, and they may play important roles in asymmetrical genome evolution in polyploidy.

Keywords: polyploidization, gene expression dominance, asymmetrical evolution, *Brassica napus*, transmitted variation

INTRODUCTION

Polyploidy is an important part of the history of natural plant species, and has promoted the domestication of cultivated crop species (Chen et al., 2018). Genome sequencing and comparative genomic analysis have identified ~50 polyploidization events across plant phylogenetic trees to date (Vanneste et al., 2014; Van de Peer et al., 2017; Chen et al., 2018). Polyploidization is considered the primary driver of plant species diversification and plays an important role in plant genome evolution. Polyploidy has become a popular research topic in plant science due to the increased number of polyploidization events identified in plants. Subgenome dominance, defined as gene fractionation

bias and expression dominance between homoeologous genes from different subgenomes, is ubiquitous in polyploidy, especially in allopolyploids. The features of transposable elements (TEs) (Freeling et al., 2012; Bottaniet al., 2018), DNA methylation (Hollister and Gaut 2009), gene density (Cheng et al., 2012), alternative splicing forms of genes (Liu et al., 2014), and single nucleotide polymorphism (SNP)-marked phenotypic diversity (Renny-Byfield et al., 2017) are substantially associated with subgenome dominance. Although understanding the mechanism of subgenome dominance has progressed, further investigation is needed to fully understand the subgenome dominance and polyploidy evolution.

Asymmetric evolution has been confirmed in many plant species since it was originally identified in the mesopolyploid *Brassica oleracea* (Liu et al., 2014). Genes in the dominant subgenome (maize 1) of maize accounted for more trait variations than those in maize 2, indicating asymmetric contributions of different subgenomes (Renny-Byfield et al., 2017). A comparison of the distribution of positively selected genes and fiber-related quantitative trait loci revealed that the A_t subgenome was selected for fiber improvement genes, and the D_t subgenome was selected for stress tolerance genes (Zhang et al., 2015b). Asymmetrical changes in wheat gene expression, small RNAs, and chromatin in subgenomes have been identified in resynthesized wheat allotetraploids (Jiao et al., 2018). *B. napus,* a recent allotetraploid species derived from *B. rapa* (AA, 2n = 20) (Wang et al., 2011) and *B. oleracea* (CC, 2n = 18) (Liu et al., 2014), is ideal for studying the effects of SNPs in asymmetric subgenomes. Investigations into asymmetric subgenomes have progressed considerably, and the gaining knowledge has played an important role in guiding genetic improvements in *B. napus.* C_n has large chromosome regions associated with important seed quality traits of rapeseed and less diversity than A_n (Qian et al., 2014; Tang et al., 2019). A_n contains a higher density of SNPs and insertion/deletions (InDels) than C_n of *B. napus*, and is consistent with a faster linkage disequilibrium (LD) decay rate of A_n than that of C_n (Wu et al., 2019). Expression levels of the genes in the A_n were significantly higher than those in the C_n subgenome, besides, A_n possesses a higher level of active epigenetic marks and lower level of repressive epigenetic marks (Zhang et al., 2021). A_n had a more homoeolog expression bias and expression level dominance than those of C_n in resynthesized *B. napus* (Wu et al., 2018). Currently, asymmetric subgenomes and asymmetric evolution have become a hotspot in polyploid research and a focus of plant science research. However, the effects of SNPs transmitted from DNA to RNA during asymmetric evolution have not been reported.

Gene expression is the process through which information from a gene is used to synthesize a functional gene that enables the generation of end products, proteins, or non-coding RNAs, ultimately affecting the phenotype (Baig et al., 2016). Mutations in SNPs are caused by environmental factors and spontaneous mutations in DNA or RNA in nature (Kato et al., 2001). Gene expression levels are affected by many factors, such as SNP density, TE density, and methylation levels (Hollister and Gaut 2009). An enhanced methylation state of TEs might suppress the expression of neighboring genes in autotetraploid rice, suggesting that chromosome doubling induces methylation variation in TEs that affect gene expression (Zhang et al., 2015a). The methylation status of TEs near genes likely plays a role in the biased expression of genes depending on the subgenome location (Hollister and Gaut 2009; Springer et al., 2016). Decreased expression of genes in a submissive subgenome in the mesopolyploid crop species *B. rapa* is associated with higher transposon coverage of DNA upstream of the transcriptional unit and increased smRNA coverage of the same region (Woodhouse et al., 2014). Genes and TEs vary among subgenomes and they are associated with gene expression in *B. rapa* (Murat et al., 2014; Cheng et al., 2016). Regulatory SNPs in genes can modify gene expression levels and affect related phenotypes (Munkhtulga et al., 2010). However, little is known about correlations between SNP variations during transcription and gene expression.

Here, we investigated the asymmetric phenomenon of SNPs transmitted from DNA to RNA in two subgenomes of *B. napus* based on genomic re-sequencing data and RNA-seq data. We assessed the relationships between nucleotide variations in genomes/transcriptomes and corresponding genes expression levels and found that they showed negative correlation. Further investigation showed that transmitted SNPs had positive correlation with gene expression level, which maybe the reason why A_n had a higher SNP density and more abundant gene expression than those of C_n in *B. napus*. Our study provides novel insights with respect to how evolvability of gene expression bias between subgenomes can be boosted by transmitted SNPs in polyploidy.

MATERIALS AND METHODS

Data Collection and SNP Identification
We investigated variations in the genomic and transcriptional profiles of two *B. napus* lines, Zhongyou821 (ZY821) and Zhongshuang11 (ZS11), which had whole-genome re-sequencing and transcriptome data from two replicates of 11 tissues. These raw data were collected from the BIG Data Center under the BioProject accession codes PRJCA000376 and PRJCA001246 (Lu et al., 2019). To identify comprehensive transcription-level SNPs, we merged all RNA-seq datasets of the same sample for variation calling. Quality was controlled by using the Fastp software with default parameters (Chen et al., 2018).

A total of 6.37 and 7.56 Gb of genomic resequencing clean bases of ZY821 and ZS11 were aligned to the reference genome (*B. napus* cv. Darmor-bzh, a French winter variety) (Chalhoub et al., 2014) using BWA-MEM v0.7.15-r1140 with default parameters (Li and Durbin 2009). Mapped reads were sorted using Samtools (Li et al., 2009), and duplicate reads of the bam alignment file were marked using Picard tools (http://broadinstitute.github.io/picard). We then used the HaplotypeCaller module of the Genome Analysis Toolkit (GATK) to identify genetic variations in DNA with default parameters (McKenna et al., 2010). Variation calling and genotyping were performed across the *Darmorbzh* reference genome. We extracted SNPs using the SelectVariants module

and filtered them using the VariantFiltration module of GATK with the following parameters: clusterSize, 3; clusterWindowSize, 10; filterExpression, QUAL <30.0 || MQ < 50.0 || QD < 2.

Transcriptional variations were identified by using 111.19 and 107.85 Gb RNA-seq clean bases of ZY821 and ZS11, and variants were called in RNA-seq data using a modified pipeline (https://github.com/gatk-workflows/gatk3-4-rnaseq-germline-snps-indels). We initially used STAR (Dobin et al., 2013) to construct index files of the reference genome and mapped clean RNA-seq reads to the reference genome with default parameters. Picard tools were then used to add ReadGroup and sort the mapped reads with "AddOrReplaceReadGroups SO = coordinate RGLB = mRNA." Then, duplicate reads were marked using the MarkDuplicates module with the parameter settings with "CREATE_INDEX = true VALIDATION_STRINGENCY = SILENT." We eliminated Ns using the SplitNCigarReads module in GATK to maintain grouping information with the parameters of rf as ReassignOneMappingQuality, RMQF 255, RMQT 60, and -U ALLOW_N_CIGAR_READS. We then identified transcriptional variations using the parameters of "-dontUseSoftClippedBases-stand_call_conf 20" in HaplotypeCaller. Finally, transcription-level SNPs were extracted and filtered by GATK using the same module and parameters as the genomic SNPs. All SNP datasets were annotated using SnpEff v4.1 (Cingolani et al., 2014) with upstream and downstream interval lengths set to 0.

Statistical Analyses of Asymmetrical SNP Transmission During DNA Transcription

To investigate the differences between the two subgenomes of *B. napus*, we compared SNP transmission between ZY821 and ZS11. Two samples or two levels containing a common homozygous SNP were regarded as having been transmitted. We were initially concerned with the transmission of genomic SNPs at the transcriptional level in ZY821 and ZS11. Because ZY821 is one of parents of ZS11, some ZY821 genetic components were transferred to the ZS11 genome during breeding. We further investigated the subgenome differences in SNPs transmitted from the ancestral parent to the offspring. We analyzed two types of transmission from ZY821 to ZS11, they were genomic SNP of ZY821 transferred to ZS11 genomic level and ZY821 genomic SNP transferred to ZS11 genomic level then transferred to ZS11 transcriptional level, respectively. The transmission ratio/rate was defined as the number of transmitted SNPs divided by the total number of genomic SNPs detected in the gene. In order to study the asymmetric distribution of transmitted SNPs in two subgenomes of *B. napus*, we analyzed the transmitted SNPs of all syntenic genes between A_n and C_n subgenomes (Chalhoub et al., 2014). Ratios of transmitted SNPs on A_n and C_n in different samples/levels were compared using two-tailed t-tests. A homozygous variant that arises from DNA to RNA transmission is regarded as a novel SNP variation in DNA to RNA.

Gene Ontology Enrichment Analyses

Pathways and biological functions of genes without SNPs and genes with genomic SNPs that were all transmitted to the transcriptional level were explored via GO enrichment analyses using the PlantRegMap database (http://plantregmap. gao-lab.org/go.php) with the parameter settings of species, *B. napus*; aspects, biological process, molecular function, cellular component; threshold, and q ≤ 0.01. The GO enrichment results are shown on the ImageGP website (http://www.ehbio.com/ImageGP/).

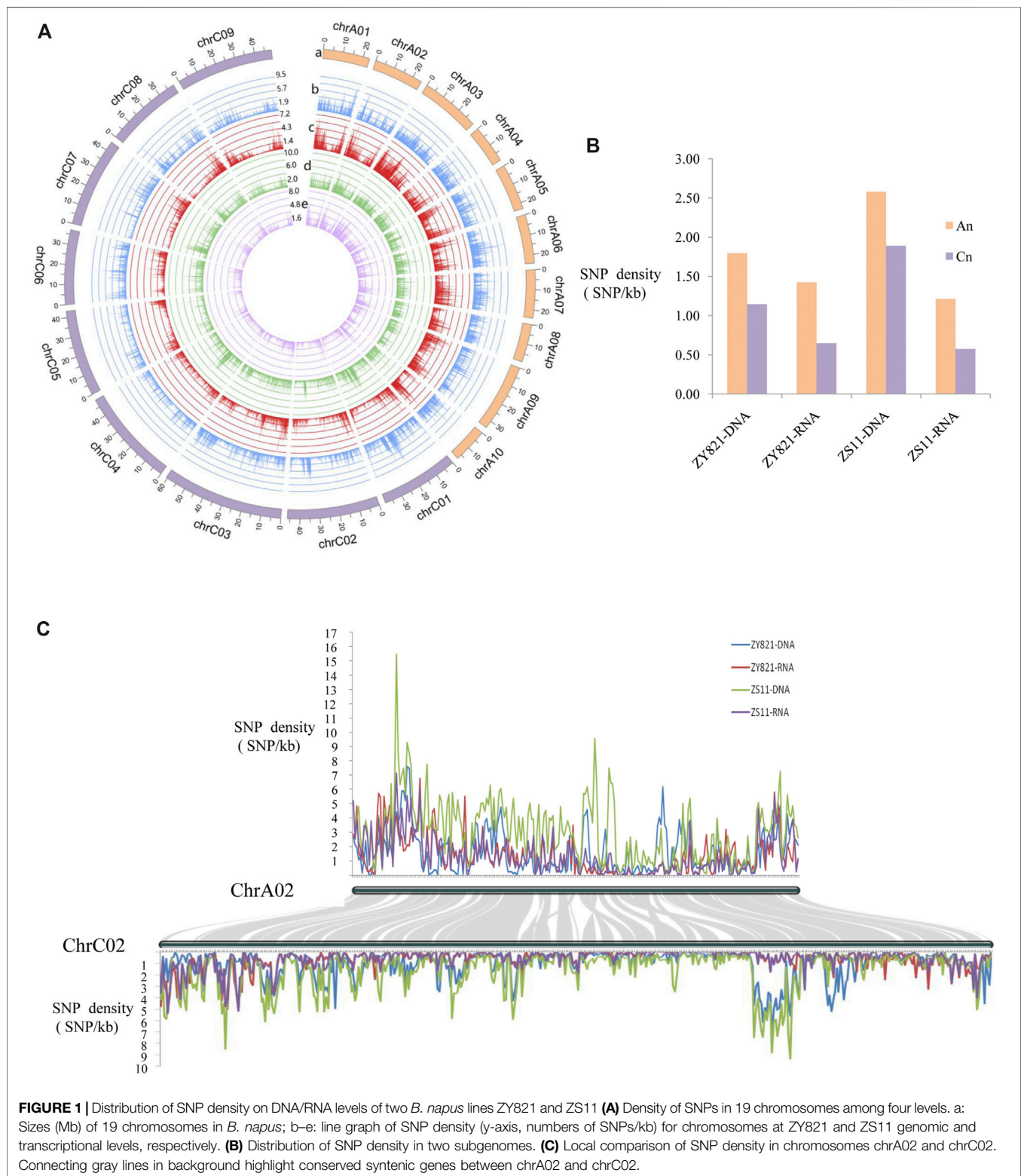## Gene Expression and Linear Regression Analyses

Filtered high-quality clean RNA-seq data from different tissues and periods of ZY821 and ZS11 were mapped to the "*Darmor-bzh*" genome (Chalhoub et al., 2014) using STAR with default parameters (Dobin et al., 2013). The mapped reads with mapping quality (MQ) ≤ 30 were filtered using Samtools, and bam files were sorted. As TPM (transcripts per kilobase of exon model per million mapped reads) could correct the inconsistences while comparing the RNA-seq abundance among independent samples (Wagner et al., 2012), StringTie (Pertea et al., 2015) was used to count unique and normalized mapped reads as TPM for each gene with the parameter of "-B -e -G".

We used linear regression analyses to investigate correlations between gene expression level and SNP numbers in gene regions, new SNP variations in DNA to RNA, and SNP transmission rates, respectively. Genes with different transmission rates were divided into 10 groups at 0.1 intervals. Genes with different numbers of SNP variations were divided into groups of 10 intervals in ZY821 DNA, ZY821 RNA, and ZS11 RNA, whereas those in ZS11 RNA were arranged into groups with 20 intervals since SNP variations occurred frequently. The linear model is $y_i = ax_i + b$, where y_i is the average TPM of the ith group, and x_i is a different group.

RESULTS

Identification and Annotation of SNPs in ZY821 and ZS11

We filtered 6.37 Gb of genome re-sequencing, and 111.19 Gb RNA-seq data were used to identify SNP variation in ZY821. In total, 1,178,526 SNPs were detected in ZY821 genomic DNA, along with 19 chromosomes and 21 scaffolds (**Supplementary Table S1**). We found that 83,179 SNPs in chromosome chrC03 had the most mutations. Among these SNPs, there were 349,509 SNPs distributed in 40,903 genic regions and 846,691 SNPs located in intergenic regions. 162,124 of the genic SNPs were located in exon regions and 139,244 SNPs in intron regions (**Supplementary Table S2**). Among the SNPs in exon regions, 69,526 missense and 91,469 synonymous variations were annotated. We also identified 792,189 SNPs at the transcription level in ZY821, and most variations were detected in chromosome chrA03 (63,925 SNPs). We found 536,701 and 144,456 variations in the exon and intron regions, respectively, and 64.05% (349,110) and 34.63% (185,846) of SNPs

FIGURE 1 | Distribution of SNP density on DNA/RNA levels of two *B. napus* lines ZY821 and ZS11 **(A)** Density of SNPs in 19 chromosomes among four levels. a: Sizes (Mb) of 19 chromosomes in *B. napus*; b–e: line graph of SNP density (y-axis, numbers of SNPs/kb) for chromosomes at ZY821 and ZS11 genomic and transcriptional levels, respectively. **(B)** Distribution of SNP density in two subgenomes. **(C)** Local comparison of SNP density in chromosomes chrA02 and chrC02. Connecting gray lines in background highlight conserved syntenic genes between chrA02 and chrC02.

in exon regions were annotated as synonymous and missense variants, respectively (**Supplementary Table S2**).

After quality control, we identified 1,823,207 and 686,858 genomic and transcriptional SNPs, respectively, in ZS11 using 7.56 Gb genome re-sequencing and 107.85 Gb RNA-seq data.

Among the genomic SNPs, 1,352,662 were located in intergenic regions with the most variations, and 228,122 and 199,500 were located in exons and introns, respectively. Overall, 97,806 (42.87%) and 128,632 (56.39%) SNPs in exons were annotated as missense and synonymous variants, respectively. Among the

SNPs at the transcription level, 464,923 and 126,513 variations occurred in the exon and intron regions, respectively. Among the SNPs in exons, 163,198 (35.10%) and 300,178 (64.57%) were missense and synonymous variants, respectively (**Supplementary Table S2**).

We identified 27,024 shared genes without SNPs at the genomic and transcriptional levels in both ZY821 and ZS11, among which 8,756 and 18,268 were located on A_n and C_n, respectively. Results of GO enrichment analysis of the genes without SNPs showed that 54 terms were enriched in cellular components, biological processes, and molecular functions. Almost all terms were basic biological processes for growth, such as an integral component of membrane and membrane part in cellular component, the regulation of nucleic acid-templated transcription and nucleoside bisphosphate metabolic processes in biological processes, enzyme inhibitor activity, and electron carrier activity in molecular function (**Supplementary Table S3**), which indicated that these genes were conserved.

Asymmetric Distribution of SNPs Between A_n- and C_n-Subgenomes

The asymmetric characteristics of the A_n and C_n subgenomes of *B. napus* have been identified (Qian et al., 2014; Lu et al., 2019; Wu et al., 2019). We found that ZY821 and ZS11 had more total SNPs in C_n than in A_n at the genomic level, whereas C_n had fewer SNPs than A_n at the transcriptional level (**Supplementary Table S1**). The total length of A_n was 315.05 Mb in the reference genome, which was less than that in the C_n (526.93 Mb) subgenome; therefore, we assessed SNP density to reevaluate differences between the two subgenomes. **Figure 1A** shows the distribution of SNP densities in the whole genome. The density of ZS11 genomic SNPs was the highest at 3.7/kb in scaffold A10_random and that of ZS11 transcriptional SNPs was the lowest (0.18/kb) in the scaffold Unn_random. These might have been due to the frequency of coding genes being the lowest (1/15.61 kb) in the scaffold Unn_random. The mean SNP density in the genome was 1.39 and 0.93/kb at the ZY821 genomic and transcriptional levels, and 2.14, and 0.81/kb at the ZS11 genomic and transcriptional levels (**Supplementary Table S1**). The genomic and transcriptional SNP density was higher in A_n than in C_n, especially at the transcriptional level, and the SNP density of A_n was almost twice that of C_n (**Figure 1B**). Taken chromosomes chrA02 and chrC02 as an example due to they had good collinearity, the SNP density of chrA02 was generally higher than that of chrC02, although that of chrA02 was occasionally lower than that of chrC02 in some chromosomal regions (**Figure 1C**).

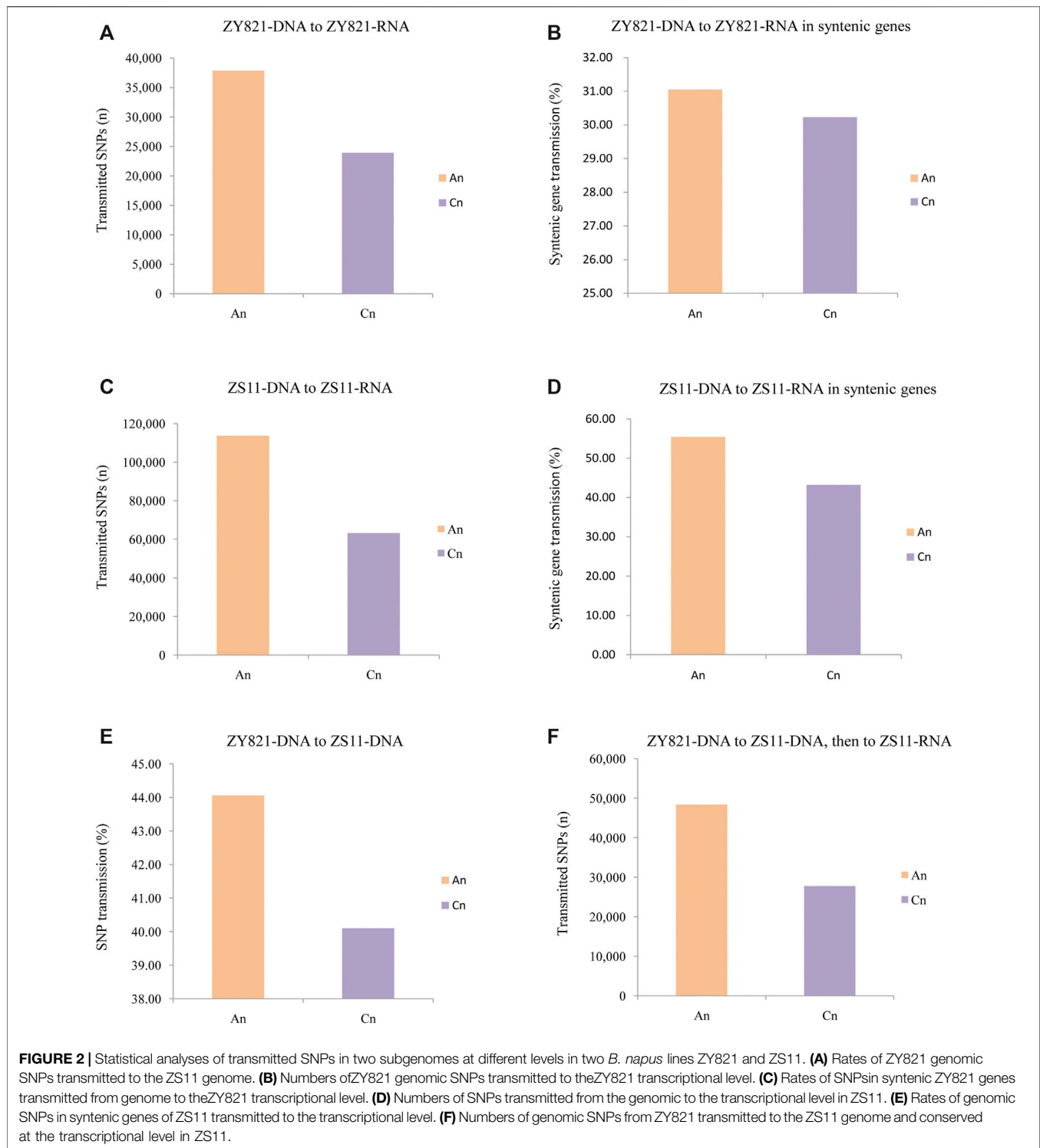## Asymmetric SNP Transmission Between A_n- and C_n-Subgenomes

We initially detected 61,832 genomic SNPs that were transmitted at the transcriptional level in ZY821 and involved 25,583 coding genes. A total of 37,845 and 23,920 SNPs occurred in A_n and C_n, respectively (**Figure 2A**), whereas 14,258 synonymous SNPs in A_n while 6,952 synonymous SNPs in C_n. Among syntenic genes in A_n and C_n, the SNP transmission ratio from DNA to RNA was 31.05% in A_n and 30.23% in C_n (**Figure 2B**). In ZS11, 113,752 SNPs were

transmitted from DNA to RNA in A_n, and 63,245 in C_n included 35,774 coding genes (**Figure 2C**). The ratios of transmitted syntenic genes in A_n and C_n were 55.43 vs. 43.21% ($p = 5.45 \times 10^{-89}$; **Figure 2D**). Parental-related transmitted SNPs differed between the two subgenomes. Among 492,916 genomic SNPs from ZY821 transmitted to ZS11, in total 249,185, 241,871, and 1,860 were in A_n, C_n, and the scaffold Unn_random, respectively. These findings indicated that 44.06% of genomic SNP in ZY821 was transmitted to the A_n of ZS11, and 40.10% of SNPs in C_n of ZY821 were transmitted to C_n of ZS11 (**Figure 2E**). Moreover, 76,188 SNPs detected at the ZY821 genomic level were transmitted to the ZS11 genome, followed by the ZS11 transcriptional level; of these, 48,367 (63.48%) of 76,188 occurred in A_n and 27,751 (36.42%) in C_n, respectively (**Figure 2F**). These results indicated that SNPs in A_n were more prone to transmission than those in C_n, especially transmission SNPs from the genomic level transfer to the transcriptional level.

The results of GO enrichment analyses revealed the functions of genes with a high ratio of SNP transmission. All (100%) genomic SNPs in 3,967 genes were transmitted to the transcriptional level in ZY821. The number of transmitted genes did not substantially differ between A_n and C_n (1,987 vs. 1,971). Only three GO terms were enriched in the biological process, viz. response to endogenous stimulus, response to hormone, and response to organic substances (**Supplementary Table S4**). Excluding biological processes, no GO terms were enriched in molecular function and cellular components (q < 0.01). Among 9,018 completely transmitted genes detected in ZS11, 5,171 and 3,835 were located in A_n and C_n, respectively. Eight GO terms were enriched in biological processes, four in cellular components, and two in molecular functions (**Figure 3**; **Supplementary Table S5**), which are involved in fundamental biological processes and stress responses. All three GO-enriched terms in the biological process of ZY821 were also enriched in ZS11, indicating that these genes play essential roles in the speciation of ZS11.
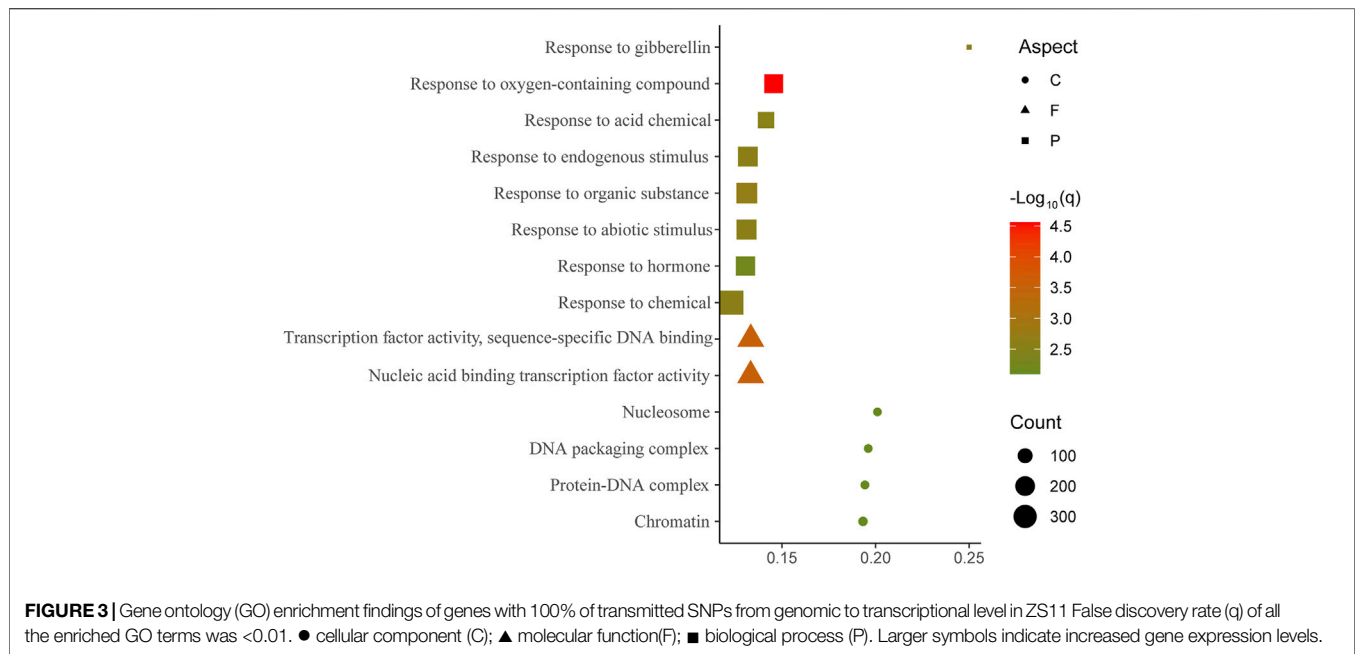
Impact of Transmitted SNPs on Asymmetrical Gene Expression

We analyzed whole gene expression levels of ZY821 and ZS11 to determine correlations between rates of SNP transmission from DNA to RNA and gene expression. We merged RNA-seq data from different tissues and periods and found that 14,717 and 15,822 genes were not expressed in ZY821 and ZS11, respectively, and 10,931 genes of them were shared, which might be because they were tissue- and time-specific. Gene expression levels decreased as the number of genomic SNPs in genes increased (**Figures 4A,B**), indicating that the expression of a gene with more SNPs would be lower in the *B. napus* genome. The expression of genes containing more RNA variations was lower in ZY821 and ZS11 during transcription (**Figures 4C,D**), which was consistent with the number of variations at the genomic and gene expression levels. However, SNP **Figures 4E,F** density was higher in A_n (**Figure 1B**), but gene expression was higher in C_n (**Figures 4E,F**). Further results revealed that transmitted SNPs had significantly associated with gene expression level. We found that 100% of 3,967 and 9,018 genes with SNPs from the genomic level were transmitted to the transcriptional level in ZY821 and ZS11,

FIGURE 2 | Statistical analyses of transmitted SNPs in two subgenomes at different levels in two *B. napus* lines ZY821 and ZS11. **(A)** Rates of ZY821 genomic SNPs transmitted to the ZS11 genome. **(B)** Numbers of ZY821 genomic SNPs transmitted to the ZY821 transcriptional level. **(C)** Rates of SNPs in syntenic ZY821 genes transmitted from genome to the ZY821 transcriptional level. **(D)** Numbers of SNPs transmitted from the genomic to the transcriptional level in ZS11. **(E)** Rates of genomic SNPs in syntenic genes of ZS11 transmitted to the transcriptional level. **(F)** Numbers of genomic SNPs from ZY821 transmitted to the ZS11 genome and conserved at the transcriptional level in ZS11.

respectively. We established a linear regression model to fit the details of the SNP transmission ratio from DNA to RNA and gene expression. SNP transmission rates were significantly and positively associated with the gene expression levels. The R^2 values of the fitted line were 0.689 and 0.735 in ZY821 and ZS11, respectively (**Figures 4G,H**), indicating that gene expression increased as the SNP transmission ratio from the DNA to the RNA level increased. These results showed that the ratio of SNPs transmitted from DNA to RNA correlated well with gene expression. The rates and numbers of transmitted SNPs were higher in A_n than in C_n (**Figure 2**), signifying asymmetric gene expression in polyploid *B. napus*.

FIGURE 3 | Gene ontology (GO) enrichment findings of genes with 100% of transmitted SNPs from genomic to transcriptional level in ZS11 False discovery rate (q) of all the enriched GO terms was <0.01. ● cellular component (C); ▲ molecular function(F); ■ biological process (P). Larger symbols indicate increased gene expression levels.

DISCUSSION

Polyploidization is considered the primary driver of plant species diversification and plays an important role in plant genome evolution, adaptive selection, gene function innovation, and crop domestication (Soltis et al., 2016). Polyploidy is a state in which the cells of an organism contain two or more sets of ancestral genomes, which may not be accompanied by multiple complete sets. Therefore, redundant genes or components in polyploidy might mutate or recombine into different neomorphic mutants with a reduced risk of extinction. Fractionation, subfunctionalization, and neofunctionalization are significant outcomes for redundant genes. Subgenome dominance develops when a subgenome expresses more genes than another when coexistent within cells (Cheng et al., 2016). The rapid development of genome sequencing and analysis technologies has led to more sequenced genomes and identified polyploidization events that have facilitated comparative genomic studies on polyploid evolution. In polyploidy *B. napus*, A_n contains a higher density of coding genes, transposable elements, and more diversity than C_n (Chalhoub et al., 2014; Wu et al., 2019). We showed a higher SNP density and more transmitted SNPs from DNA to DNA/RNA in A_n, especially from the DNA level to the RNA level, which complements the asymmetric feature in polyploidy. The asymmetric features in the two subgenomes might be due to more frequent outcrossing between *B. napus* and *B. rapa* than between *B. napus* and *B. oleracea* in terms of breeding behaviors (Wu et al., 2019). The A_n subgenome of a modern *B. napus* could retain excellent variations in artificial selection according to the aims of breeders. The GO findings showed that genes with completely transmitted SNPs of ZY821 and ZS11 were commonly enriched in fundamental biological processes and stress responses. This might be an excellent stress SNP micromutation in coding genes

that facilitate disease resistance. Therefore, they have been selected by breeders considering adaptability and disease resistance and they were widely planted in China during recent decades.

Gene expression is the most fundamental level at which a genotype produces observable traits. The regulation of gene expression is the basis for cellular differentiation, development, morphogenesis, versatility, and adaptability of all organisms. The amount of gene expression level was affected by SNP density, TE density, and methylation level. Regulatory factors containing a splice acceptor and promoter in intergenic regions are important for regulating gene expression (LeBowitz et al., 1993; Mette et al., 2000; Weber et al., 2007). Different profiles of methylation and altered splicing forms are also associated with gene expression levels (Liu et al., 2014; Mei et al., 2017; Renny-Byfield et al., 2017). Transposons can be repressed, and the expression of genes encoding proteins can be affected by RNA-directed DNA methylation (Rowley et al., 2017). The present findings showed that the numbers and transmission rates of SNPs were associated with gene expression. Changes in SNP bases might result in different combinations of transcription factors, which would lead to different levels of coding gene expression. Transposable elements, splice forms, methylation modifications, and SNP density contribute to gene expression, but systematic understanding remains limited. The integration of genomics, transcriptomics, epigenetics, morphology, and synthetic methods in appropriate species is essential for gaining further insight.

Transcription factors are proteins that control the rates at which genetic information is transcribed from DNA to messenger RNA by binding to a specific DNA sequence, and they are essential for regulating gene expression. Mutations in transcription factors may affect the expression of related genes. A specific binding SNP in the first intron of the *SPP1*
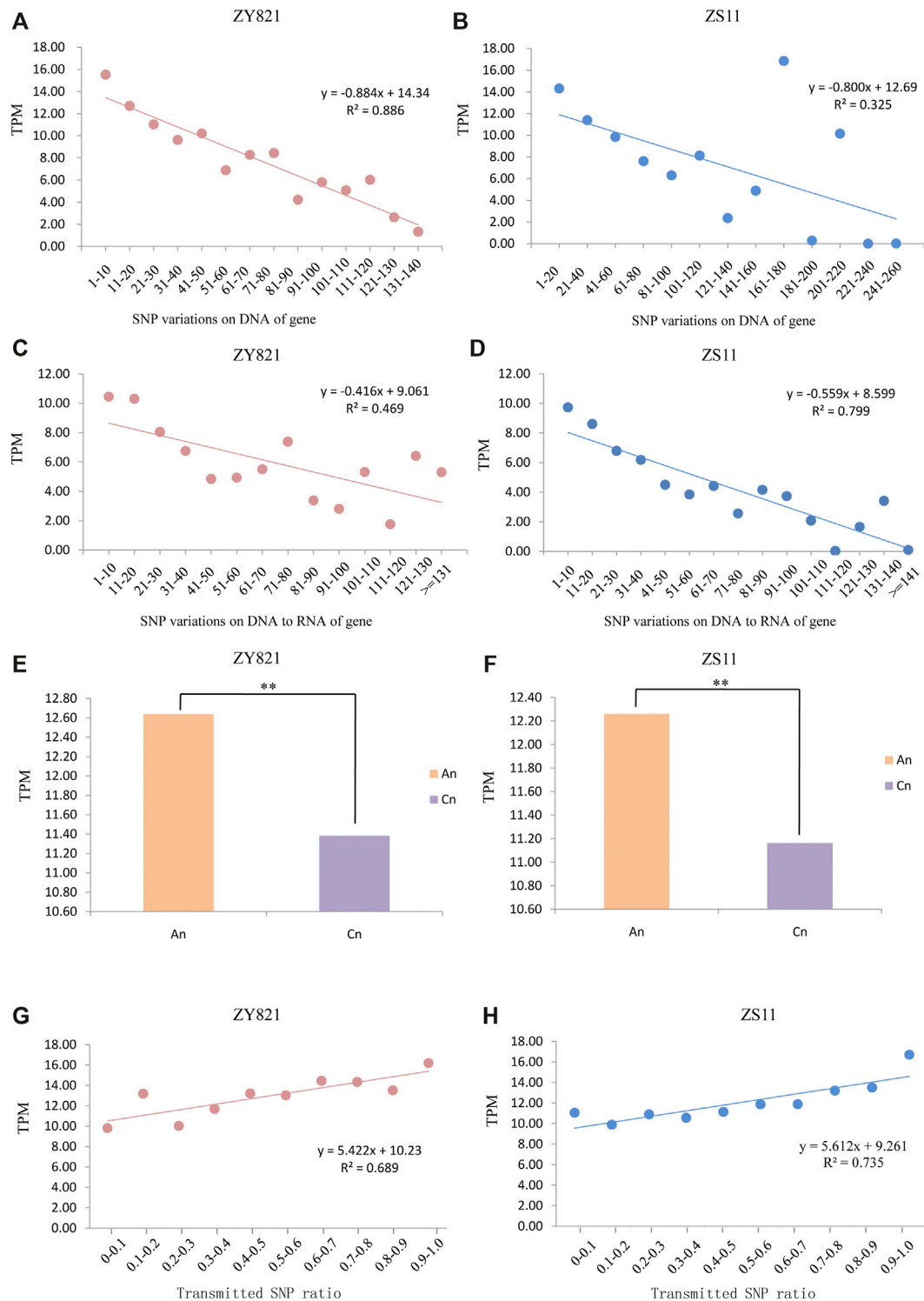
FIGURE 4 | Correlations of the gene expression levels with SNP variations and transmission ratio Correlations and linear regression between gene expression (TPM) and numbers of genomic SNP in all ZY821 **(A)** and ZS11 **(B)** genes. Correlations and linear regression analyzes between TPM and numbers of novel SNP variations on DNA to RNA of gene in ZY821 **(C)** and ZS11 **(D)**. Average expression levels of all syntenic genes in A_n and C_n in ZY821 **(E)** and ZS11 **(F)**, ** represent $p < 0.01$. Correlations and linear regression analyzes between TPM and ratios of SNPs of all genes transmitted from DNA to RNA in ZY821 **(G)** and ZS11 **(H)**.

gene affects the functional characteristics of *SPP1* at the DNA and RNA levels by activating abnormal splicing of the first intron (Muráni et al., 2009). A changed base in rs3122605-G upregulated *IL10* gene expression and increased the risk of systemic lupus erythematosus in European Americans (Sakurai et al., 2013). Our findings showed that gene expression decreased with increasing DNA (or DNA to RNA) variations in coding genes. This indicated that more variations at the level of DNA and during transcription from DNA to RNA resulted in lower gene expression. We excluded high-density variations that occur due to fewer mismatch mapping transcriptional reads, and determined the SNP frequency within 150 bp windows (the read length of RNA-seq in this study) in the entire genome (5,668,614 windows). We found that only 68 and 38 windows in ZY821 and ZS11 contained more than 10 SNPs (default mismatch mapping parameter of STAR), indicating that genes with a higher density of RNA variation did not occur due to a few mismatch reads. This may be because more genomic SNP variations in the gene would increase the risk of a transcription factor not binding to a specific DNA sequence, and nucleotide variations during the transcription process from DNA to RNA might affect RNA alternative splicing and transcriptional inhibition. Alternative splicing of precursor mRNAs from multiexon genes allows an organism to increase its coding potential and regulate gene expression through multiple mechanisms (Reddy et al., 2013). We found 19,756 and 17,509 RNA SNPs in alternative splice regions in ZY821 and ZS11, respectively. These SNPs may affect gene expression via alternative splicing.

Subgenome expression dominance might be obscured by potential homoeologous exchange between different subgenomes, for example, in wheat (Pfeifer et al., 2014; Harper et al., 2016) and *B. napus* (Chalhoub et al., 2014). We found lower expression of genes with more SNP variations, but A_n had not only a higher SNP density, but also a higher level of gene expression. Further analyses revealed that the transmitted SNP ratio was positively correlated with gene expression level. The SNPs transmitted from DNA to RNA might be excellent variations that facilitated the formation of traits that were favored and selected by breeders, and finally retained. Genome dominance is heritable; the expression of a gene in a dominant subgenome tends to be higher than that of its recessive homeolog (Woodhouse et al., 2014). Transmitted SNPs vary in homologous genes in the corresponding subgenome and might be associated with gene dominance. Given our understanding of the subgenome dominance of *B. napus*, the selected SNPs transmitted from DNA to RNA played important roles in gene expression, but the random SNPs involved in down-regulating gene expression might obscure gene expression dominance. In this study, we used the statistical methods to discover a positive correlation between transmitted SNPs and gene expression level, besides this, the effect of transmitted SNPs on gene expression level and subgenome dominance needs to be verified in more polyploidy species and direct molecular experiment evidence. Asymmetric divergence and genome evolution are popular topics in plant science, especially in polyploidy. The innovative knowledge of transmitted SNPs might be an important field in genome evolution of polyploid crops.

CONCLUSION

We identified variations in genomic and transcriptional SNPs in ZY821 and ZS11 and investigated the correlation between SNP numbers/transmission ratios and gene expression. We showed that variations in DNA or the transcription process from DNA to RNA correlated negatively with gene expression levels. The extent of gene expression increased with an increasing ratio of SNP transmission from DNA to RNA, which might explain why A_n had a higher SNP density and more abundant gene expression. SNPs transmitted from DNA to DNA or DNA to RNA might play important roles in asymmetrical evolution and subgenome dominance in polyploidy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YZ and JC designed the study and revised the manuscript. MT carried out the study and wrote the manuscript. JL revised the manuscript. LS assisted in writing the manuscript. MH improved the language of the manuscript. XH drew the figures. All authors contributed to the manuscript and approved the final manuscript to publish.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.756172/full#supplementary-material

REFERENCES

Baig, M., Keservani, R., and Sharma, A. K. (2016). *RNA Sequencing and Gene Expression Regulation*. Chapter. Boca Raton: Computational Biology and Bioinformatics[M], 71–105.

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

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., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Cheng, F., Sun, C., Wu, J., Schnable, J., Woodhouse, M. R., Liang, J., et al. (2016). Epigenetic Regulation of Subgenome Dominance Following Whole Genome Triplication in Brassica Rapa. *New Phytol.* 211, 288–299. doi:10.1111/nph.13884

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

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., et al. (2012). Biased Gene Fractionation and Dominant Gene Expression Among the Subgenomes of Brassica Rapa. *PLoS One* 7, e36442. doi:10.1371/journal.pone.0036442

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2014). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* 6, 80–92. doi:10.4161/fly.19695

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

Freeling, M., Woodhouse, M. R., Subramaniam, S., Turco, G., Lisch, D., and Schnable, J. C. (2012). Fractionation Mutagenesis and Similar Consequences of Mechanisms Removing Dispensable or Less-Expressed DNA in Plants. *Curr. Opin. Plant Biol.* 15, 131–139. doi:10.1016/j.pbi.2012.01.015

Harper, A. L., Trick, M., He, Z., Clissold, L., Fellgett, A., Griffiths, S., et al. (2016). Genome Distribution of Differential Homoeologue Contributions to Leaf Gene Expression in Bread Wheat. *Plant Biotechnol. J.* 14, 1207–1214. doi:10.1111/pbi.12486

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

Jiao, W., Yuan, J., Jiang, S., Liu, Y., Wang, L., Liu, M., et al. (2018). Asymmetrical Changes of Gene Expression, Small RNA S and Chromatin in Two Resynthesized Wheat Allotetraploids. *Plant J.* 93, 828–842. doi:10.1111/tpj.13805

Kato, J., Fujikawa, K., Kanda, M., Fukuda, N., Sasaki, K., Takayama, T., et al. (2001). A Mutation, in the Iron-Responsive Element of H Ferritin mRNA, Causing Autosomal Dominant Iron Overload. *Am. J. Hum. Genet.* 69, 191–197. doi:10.1086/321261

LeBowitz, J. H., Smith, H. Q., Rusche, L., and Beverley, S. M. (1993). Coupling of Poly(A) Site Selection and Trans-splicing in Leishmania. *Genes Dev.* 7, 996–1007. doi:10.1101/gad.7.6.996

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

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., et al. (2014). The *Brassica oleracea* Genome Reveals the Asymmetrical Evolution of Polyploid Genomes. *Nat. Commun.* 5, 3930. doi:10.1038/ncomms4930

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome Resequencing Reveals Brassica Napus Origin and Genetic Loci Involved in its Improvement. *Nat. Commun.* 10, 1154. doi:10.1038/s41467-019-09134-9

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Mei, W., Liu, S., Schnable, J. C., Yeh, C.-T., Springer, N. M., Schnable, P. S., et al. (2017). A Comprehensive Analysis of Alternative Splicing in Paleopolyploid Maize. *Front. Plant Sci.* 8, 694. doi:10.3389/fpls.2017.00694

Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A., and Matzke, A. J. M. (2000). Transcriptional Silencing and Promoter Methylation Triggered by Double-Stranded RNA. *Embo J.* 19, 5194–5201. doi:10.1093/emboj/19.19.5194

Munkhtulga, L., Nagashima, S., Nakayama, K., Utsumi, N., Yanagisawa, Y., Gotoh, T., et al. (2010). Regulatory SNP in theRBP4Gene Modified the Expression in Adipocytes and Associated with BMI. *Obesity* 18, 1006–1014. doi:10.1038/oby.2009.358

Muráni, E., Ponsuksili, S., Seyfert, H.-M., Shi, X., and Wimmers, K. (2009). Dual Effect of a Single Nucleotide Polymorphism in the First Intron of the Porcine Secreted Phosphoprotein 1 Gene: Allele-specific Binding of C/EBP Beta and Activation of Aberrant Splicing. *BMC Mol. Biol* 10, 96. doi:10.1186/1471-2199-10-96

Murat, F., Zhang, R., Guizard, S., Flores, R., Armero, A., Pont, C., et al. (2014). Shared Subgenome Dominance Following Polyploidization Explains Grass Genome Evolutionary Plasticity from a Seven Protochromosome Ancestor with 16K Protogenes. *Genome Biol. Evol.* 6, 12–33. doi:10.1093/gbe/evt200

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

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

Qian, L., Qian, W., and Snowdon, R. J. (2014). Sub-genomic Selection Patterns as a Signature of Breeding in the Allopolyploid Brassica Napus Genome. *BMC Genomics* 15, 1170. doi:10.1186/1471-2164-15-1170

Reddy, A. S. N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the Alternative Splicing Landscape in Plants. *Plant Cell* 25, 3657–3683. doi:10.1105/tpc.113.117523

Renny-Byfield, S., Rodgers-Melnick, E., and Ross-Ibarra, J. (2017). Gene Fractionation and Function in the Ancient Subgenomes of Maize. *Mol. Biol. Evol.* 34, 1825–1832. doi:10.1093/molbev/msx121

Rowley, M. J., Rothi, M. H., Böhmdorfer, G., Kuciński, J., and Wierzbicki, A. T. (2017). Long-range Control of Gene Expression via RNA-Directed DNA Methylation. *Plos Genet.* 13, e1006749. doi:10.1371/journal.pgen.1006749

Sakurai, D., Zhao, J., Deng, Y., Kelly, J. A., Brown, E. E., Harley, J. B., et al. (2013). Preferential Binding to Elk-1 by SLE-Associated IL10 Risk Allele Upregulates IL10 Expression. *Plos Genet.* 9, e1003870. doi:10.1371/journal.pgen.1003870

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

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

Tang, M., Zhang, Y., Liu, Y., Tong, C., Cheng, X., Zhu, W., et al. (2019). Mapping Loci Controlling Fatty Acid Profiles, Oil and Protein Content by Genome-wide Association Study in Brassica Napus. *Crop J.* 7, 217–226. doi:10.1016/j.cj.2018.10.007

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., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 Plant Genomes Supports a Wave of Successful Genome Duplications in Association with the Cretaceous-Paleogene Boundary. *Genome Res.* 24, 1334–1347. doi:10.1101/gr.168997.113

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA Abundance Using RNA-Seq Data: RPKM Measure Is Inconsistent Among Samples. *Theor. Biosci.* 131, 281–285. doi:10.1007/s12064-012-0162-3

Wang, X., Wang, H., Wang, H., Wang, J., Sun, R., Wu, J., et al. (2011). The Genome of the Mesopolyploid Crop Species Brassica Rapa. *Nat. Genet.* 43, 1035–1039. doi:10.1038/ng.919

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., et al. (2007). Distribution, Silencing Potential and Evolutionary Impact of Promoter DNA Methylation in the Human Genome. *Nat. Genet.* 39, 457–466. doi:10.1038/ng1990

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.* 111, 5283–5288. doi:10.1073/pnas.1402475111

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., et al. (2019). Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. *Mol. Plant* 12, 30–43. doi:10.1016/j.molp.2018.11.007

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

Zhang, J., Liu, Y., Xia, E. H., Yao, Q. Y., Liu, X. D., and Gao, L. Z. (2015a). Autotetraploid rice Methylome Analysis Reveals Methylation Variation of Transposable Elements and Their Effects on Gene Expression. *Proc. Natl. Acad. Sci. U S A.* 112, E7022–E7029. doi:10.1073/pnas.1515170112

Zhang, Q., Guan, P., Zhao, L., Ma, M., Xie, L., Li, Y., et al. (2021). Asymmetric Epigenome Maps of Subgenomes Reveal Imbalanced Transcription and Distinct Evolutionary Trends in Brassica Napus. *Mol. Plant* 14, 604–619. doi:10.1016/j.molp.2020.12.020

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015b). Sequencing of Allotetraploid Cotton (Gossypium Hirsutum L. Acc. TM-1) Provides a Resource for Fiber Improvement. *Nat. Biotechnol.* 33, 531–537. doi:10.1038/nbt.3207

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.