agronomy

MDPI

Article

Epigenetic and Genetic Contribution for Expression Bias of Homologous Alleles in Polyploid Sugarcane

Yajie Xue [1,†], Yan Shi [1,†], Yuchuan Qi [1,†], Hang Yu [1], Chengwu Zou [1], Xueying Cao [1], Cuimin Sun [2], Baoshan Chen [1,*] and Haifeng Wang [1,*]

[1] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi Key Lab for Sugarcane Biology, National Demonstration Center for Experimental Plant Science Education, College of Agriculture, Guangxi University, Nanning 530004, China
[2] School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China
* Correspondence: chenbs2008@163.com (B.C.); haifengwang@gxu.edu.cn (H.W.)
† These authors contributed equally to this work.

Abstract: DNA methylation regulates gene expression in eukaryotes, but their roles in gene expression changes in polyploids are poorly understood. Here, we comparatively analyzed the DNA methylation profiles and transcriptome maps of four tissues (leaf, rind, pith, and root) in autopolyploid sugarcane (*Saccharum spontaneum*). The overall DNA methylation levels were relatively equal and were consistent with the similar genome-wide expression levels of homologous alleles among different homologous chromosomes. However, tetrad alleles showed a large proportion of non-balanced alleles (85–89%) compared with balanced alleles (11–15%). For tetrad alleles, highly-expressed alleles showed lower *Ka/Ks* values than those in lowly-expressed alleles, indicating that dominant (and non-suppressed) alleles were under strong purifying selection pressure compared to non-dominant (and suppressed) alleles. Specifically, higher DNA methylation levels were found in non-balanced alleles compared to balanced alleles, as well as a higher CG to TG substitution rate, suggesting epigenetic and genetic variation associated with transcription divergence. Moreover, among non-balanced alleles, which were associated with DNA methylation changes, many were involved in several processes relevant to agronomic traits, such as responses to stress and carbohydrate transport. Taken together, our results provide unique epigenetic insights into the transcriptional regulation of homologous alleles, which contribute to the key agronomic traits in sugarcane and facilitate epigenetic studies in other polyploid crops.

Keywords: epigenetics; DNA methylation; sugarcane; polyploid; homologous alleles

Citation: Xue, Y.; Shi, Y.; Qi, Y.; Yu, H.; Zou, C.; Cao, X.; Sun, C.; Chen, B.; Wang, H. Epigenetic and Genetic Contribution for Expression Bias of Homologous Alleles in Polyploid Sugarcane. *Agronomy* **2022**, *12*, 2852. https://doi.org/10.3390/agronomy12112852

Academic Editor: Yang Zhu

Received: 9 October 2022
Accepted: 12 November 2022
Published: 15 November 2022

1. Introduction

Sugarcane is a major crop for sugar and biofuel production, originating from Papua New Guinea. It belongs to the Poaceae family and genus *Saccharum*, which comprises six species, including three sugar-bearing species *S. officinarum*, *S. sinense*, and *S. barberi*, and three non-sugar-bearing species *S. spontaneum*, *S. robustum* and *S. edule*. The cultivated sugarcane is an allopolyploid with high ploidy level (2n $\geq$ 10x = 100–130), which contains genomes of *S. officinarum* and *S. spontaneum* [1]. The genome of sugarcane is always polyploidy, highly heterozygous, and aneuploidy. Its large genome size, polyploidy, and high repeated contents (>65%) have made genome sequencing and analysis challenging [2,3]. As a result, molecular biology and functional genomics studies lag behind far more for sugarcane than for the other important crops, including rice, maize and cotton. Recently, with the development of sequencing technologies, many chromosome-scale high-quality genomes were assembled by a combination of third-generation sequencing techniques and high-throughput chromatin conformation capture techniques (Hi-C) [4], including many polyploidy plant genomes [5,6], of which autopolyploid sugarcane (*S. spontaneum*) is one of them [2].

Whole genome duplication (WGD or polyploidy) is widespread in eukaryotes, particularly in plants, and plays an important role in speciation and environment adaptation [7], for example in cotton, wheat, maize, and soybean, which have all experienced several ancient WGD events [8,9]. Cumulative evidence has also shown that WGD plays an important role in the domestication and evolution of the agronomic traits of these species [9,10]. The polyploid produced by WGD can be divided into two types, one is allopolyploid and the other is autopolyploid. Most of studies of polyploid are focused on allopolyploids, such as cotton, peanut, wheat, etc., showing the importance of subgenome dominance, homoeolog expression bias, or genomic variation during the evolution of allopolyploids [6,11,12]. However, compared with allopolyploids, the research on autopolyploids is very limited. Sugarcane (*S. spontaneum*) is an important autopolyploid plant, which has been sequenced and assembled into four groups of homologous chromosomes. With homologous alleles identification, it provides us with an unprecedented opportunity to study the evolution and expression changes between the homologous alleles of different homologous chromosomes. Combined with the knowledge from allopolyploids and autopolyploids, it gives us a more comprehensive understanding of WGD's role in the evolution of polyploids. The dynamic DNA methylation patterns during subgenome evolution and homoeologous expression regulation have been demonstrated to be a critical regulation mechanism in polyploid plants [13–15]. Recent studies have shown that biased expression of homoeologs is correlated with changes in DNA methylation in many allopolyploids [14,15], such as *B. napus*, wheat and cotton [13,16,17]. However, we have limited understanding of the extent of homologous alleles in autopolyploids.

To explore the potential role of DNA methylation changes on the expression divergence of homologous alleles and the functional consequence in autopolyploid sugarcane, we investigated DNA methylation profiles across four groups of homologous chromosomes of sugarcane (*S. spontaneum*) genome across four tissue types. This allowed us for to explore epigenetic variation in this large complex genome for the first time and open up a new level of epigenetic variation, which can be exploited by breeders. We implemented this genome-wide approach data to address three questions: how does DNA methylation play a role in homologous alleles expression bias? Is this phenomenon consistent with findings in allopolyploid plants? What is the functional consequence of those biased expression homologous alleles regulated by changes in DNA methylation?

2. Materials and Methods

2.1. Library Construction and Sequencing

DNA was extracted from the young leaf, root, rind, and pith of sugarcane (*S. officinarum*) (Guangxi University, Nanning, China, 22.82°N, 108.37°E), and BS-seq libraries were prepared using the TruSeq Nano DNA LT kit (Illumina, Beijing, China), as described previously [18]. Two libraries were sequenced on the NovaSeq 6000 system (Illumina) to obtain paired-end 150-bp reads as per the manufacturer's instructions.

Total RNA was extracted from the same tissue as BS-seq libraries, and RNA-seq libraries were prepared using a TruSeq preparation kit with polyA mRNA selection, as per the manufacturer's instructions (Illumina). Three libraries were pooled and sequenced for paired-end 150-bp reads using an Illumina NovaSeq 6000 system.

2.2. DNA Methylation Sequencing Data Analysis

Low quality sequencing reads were filtered using FASTP with a default parameter [19], and the remaining reads were aligned to the sugarcane reference genome using BSMAP 2.8.7 [20]. Only uniquely mapped paired-end reads were used to estimate methylation ratios. The methylation ratio was calculated as the number of Cs divided by Cs plus Ts.

2.3. Transcriptome Sequencing Data Analysis

High quality sequencing reads were aligned to the sugarcane reference genome by using HiSAT2 [21]. Only uniquely mapped paired-end reads were left for further study.

Expression values were estimated by FPKM (fragments per kilobase per million reads). Homologous alleles with an average FPKM of more than 0.5 across different tissues were considered as expressed.

For tetrads expression bias categories classification, we applied a similar criteria as the one used in the reference [22]. We standardized the relative expression of each homologous allele across the tetrads as follows:

$$\text{Expression}_A = \text{FPKM}_A / (\text{FPKM}_A + \text{FPKM}_B + \text{FPKM}_C + \text{FPKM}_D)$$

$$\text{Expression}_B = \text{FPKM}_B / (\text{FPKM}_A + \text{FPKM}_B + \text{FPKM}_C + \text{FPKM}_D)$$

$$\text{Expression}_C = \text{FPKM}_C / (\text{FPKM}_A + \text{FPKM}_B + \text{FPKM}_C + \text{FPKM}_D)$$

$$\text{Expression}_D = \text{FPKM}_D / (\text{FPKM}_A + \text{FPKM}_B + \text{FPKM}_C + \text{FPKM}_D)$$

where A, B, C, and D represent the gene corresponding to the A, B, C, and D homologous alleles in tetrad.

2.4. Ka/Ks Analysis

The *Ka/Ks* ratios of the orthologous gene pairs between sugarcane and sorghum were calculated by a python script, including CLUSTALW, PAL2NAL and PAML (https://github.com/tanghaibao/bio-pipeline/tree/master/synonymous_calculation (accessed on 10 October 2020)).

2.5. GO Enrichment Analysis

The Omicshare tool (www.omicshare.com/tools (accessed on 1 March 2020)) is an online data analysis platform for hypergeometric testing, which is used for GO enrichment of differentially expressed methylation-associated alleles. Only GO terms with p value less than 0.05 were used for further analysis.

2.6. Genetic Substitution Rate Calculation

The genetic substitution rate is calculated as the proportion of different sites among the total sites that could be aligned between sorghum and sugarcane orthologous gene pairs. Orthologous gene pair alignments were generated on coding sequences with MAFFT (v7.453) [6] and a custom Perl script was implemented to calculate the genetic substitution rate.

3. Results

3.1. Expression Patterns of Homologous Alleles in Autopolyploid Sugarcane

In sugarcane, we did not reveal any genome-wide significant bias of gene content, structure and composition between the different sub-genomes [2]. To determine the expression patterns of homologous alleles, we sequenced and analyzed transcriptome data from the leaf, root, rind and pith of the sugarcane at the maturity stage. At the whole genome level, we found no overall expression bias of homologous alleles among the four groups of homologous chromosomes in the leaf, and this phenomenon was also observed across the other three tissues in the sugarcane (Figure 1A). Consistent with the results in the previous studies of allopolyploid crops, there is also no significant genome-wide expression bias between homoeologs from different subgenomes, such as cotton (*Gossypium hirsutum*) [23], rapeseed (*Brassica napus*) [24], senvy (*B. juncae*) [25], and wheat (*Triticum aestivum*) [26]. In addition, we also used RT-PCR of selected four DNA methyltransferase genes to verify the accuracy of our RNA-seq data, and we found consistent results between the RT-PCR data and RNA-seq data of those selected genes. These findings suggest our high accuracy and reproducibility of RNA-seq data (Figure S1A and Table S1).

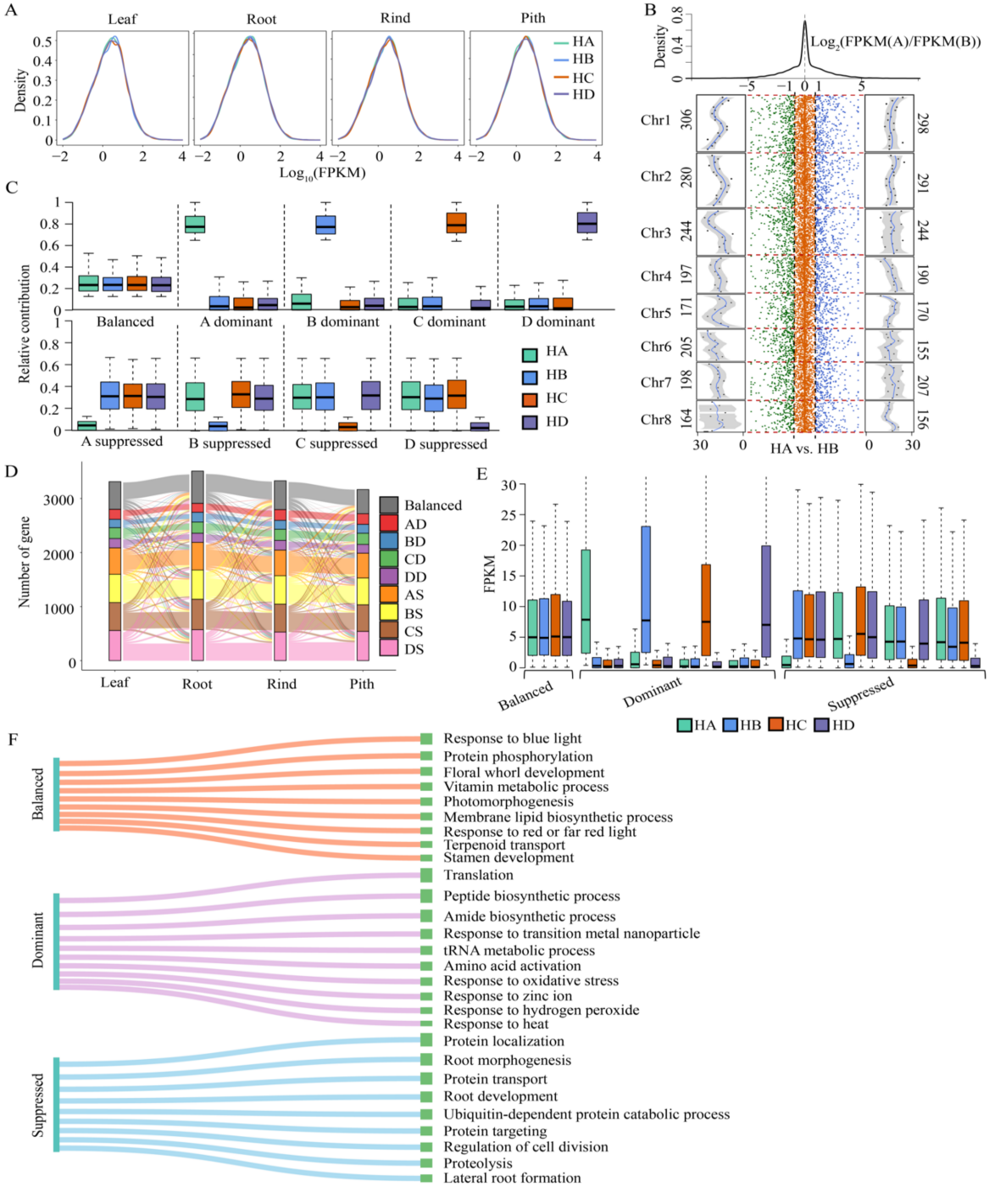

Figure 1. Expression patterns of homologous alleles in autopolyploid sugarcane. (**A**) Genome-wide gene expression distribution across different tissues of sugarcane. (**B**) Homologous alleles expression bias and location preference of comparison between HA an HB. (**C**) Nine expression categories of homologous tetrad alleles. (**D**) Proportion of nine expression categories across four tissues. (**E**) Comparison of expression level of tetrad alleles from nine categories. (**F**) Enriched GO functions of balanced, dominant and suppressed tetrad alleles.

The overall pairwise expression correlation between the subgenomes was very similar, and no evidence of individual subgenomes in the genome-wide transcription dominance was observed (Figures 1B and S1B–F). However, thousands of homologous allele pairs were differentially expressed, which means that one homologous allele is highly expressed, and the other allele is lowly expressed (fold-change > 2) (Figure 1B). For example, in

the comparison of homologous chromosomes 1 between HA (Haplotype A) and HB, we found that 298 homologous alleles were highly expressed in HA, while 306 homologous alleles were highly expressed in HB. Unlike the hexaploid wheat genome, differentially expressed homoeologous genes were overrepresented in the distal ends of the chromosomes, which have high recombination rates [22]. We found that differentially expressed homologous alleles in sugarcane have no obvious positional preference (Figure 1B). In sugarcane genome, 3977, 7713, 10,904 genes were defined with four alleles, three alleles, and two alleles, respectively [2]. Due to the complexity of homologous alleles among four sets of homologous sub-genomes, as well as to facilitate our examination of expression divergence of homologous alleles, we focused on 3977 genes with four alleles defined that had a 1:1:1:1 correspondence across the four homologous genomes, referred to as tetrads. We only retained the tetrads with the summed expression of >0.5 FPKM (95% of all tetrads) for later analysis. According to the methods mentioned in reference [22], we defined nine homologous alleles expression categories: one balanced category with similar expression levels from the four alleles; and eight allele-dominant or allele-suppressed categories (Figure 1C and Table S2). Based on the classification above, we found tetrads assigned to the balanced category were relatively rare (from 14% in the pith to 17% in the root) (Figure 1D). Tetrads with single-allele dominant were infrequent (21% to 23%), whereas single-allele suppressed were more common (61% to 63%). Strikingly, most of alleles were assigned to the same categories across different tissues. The suppressed or dominant expression pattern shared by the different tissues of the sugarcane suggests that the repressed or active effect is the result of a conserved set of homologous alleles, and there is a relatively stable factor that controls the expression level of homologous alleles among subgenomes in sugarcane. Unexpectedly, these patterns are inconsistent to those in the allopolyploid wheat genome [22]. In the wheat genome, balanced homoeologs were found to be expressed more widely and higher than dominant or suppressed homoeologs, but the higher expression levels of dominant alleles were observed in sugarcane when compared to balanced or suppressed alleles (Figure 1E). This inconsistent result may suggest different regulatory mechanisms in homologous alleles and homoeologs from different subgenomes in autopolyploid sugarcane and allopolyploid wheat, respectively. Gene ontology analysis of balanced and non-balanced categories suggested different mechanisms of homologous alleles expression bias. For example, homologous alleles of balanced categories were always involved in the response to light and flower development pathways, such as the response to blue light, the response to red or far red light, floral whorl development and stamen development (Figure 1F and Table S3), but dominant alleles were enriched in house-keeping processes (translation) and responses to external stimuli (responses to transition metal nanoparticles, responses to oxidative stress and responses to hydrogen peroxide) (Figure 1F and Table S4). In addition, suppressed alleles showed very different functional categories from balanced and dominant alleles, and protein localization, transport pathways and root development were most represented in suppressed alleles (Figure 1F, and Table S5). These results suggest that different types of homologous alleles participate in divergent regulatory processes.

3.2. Role of DNA Methylation in Homologous Alleles Expression Bias

Genome-wide homologous allele pairs did not show expression bias in multiple tissues (Figures 1A,B and S1B–F), whereas at least 80% of tetrads can be classified into the non-balanced (dominant or suppressed tetrads) group in a given tissue. To investigate the role of DNA methylation in homologous alleles expression bias in autopolyploid sugarcane, we firstly examined the DNA methylation levels of homologous alleles, and found relatively equal DNA methylation between homologous alleles (Figures 2A and S2A,D), which is consistent with the finding of homoeologs of methylation patterns in allopolyploid cotton [13]. However, pairwise comparison showed thousands of homologous alleles were differently methylated among different subgenomes (Figure 2B for CHG, Figure S2B,E for CG and CHH, respectively). Next, we divided these alleles of bias methylation into hyper-

methylated alleles and hypo-methylated alleles and compared the changes of expression. We found non-CG methylation (e.g., CHG and CHH) was anti-correlated with expression levels of homologous alleles at gene body regions (Figures 2C and S2F). However, CG methylation showed a complex relationship with expression levels of homologous alleles (Figure S2C), which is consistent with a previous study on homoeologs expression from two different subgenomes of allotetraploid cotton [13]. This analysis indicated that the mechanism of non-CG methylation in gene body regions regulating the expression divergence of homologous genes is conserved among different plants, no matter in autopolyploids or allopolyploids [27].

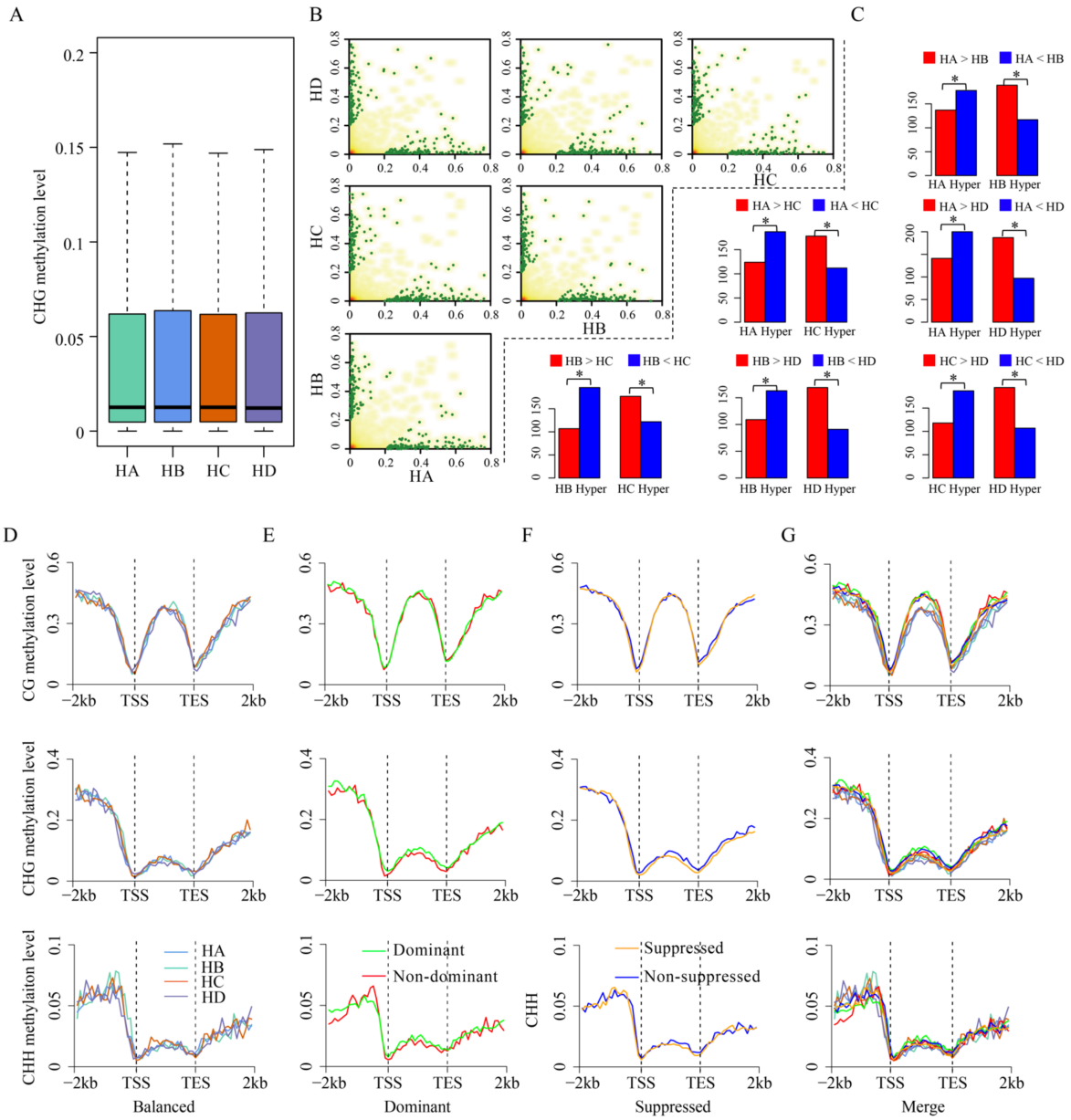

Figure 2. DNA methylation variation and association with expression of tetrad alleles. (**A**) Boxplot showing CHG methylation level of tetrad alleles from different haplotype chromosomes. (**B**) Pairwise comparison of CHG methylation level among four groups of homologous chromosomes. (**C**) Comparison of the number of homologous alleles with different gene body methylation and expression. The differences between homologous alleles were tested using the Wilcoxon rank sum test. * p value < 0.05. (**D**–**G**) DNA methylation profile of balanced (**D**), dominant (**E**), suppressed (**F**), and merged (**G**) tetrad alleles in CG, CHG, and CHH sequence contexts, respectively.

Although DNA methylation at gene body regions is widely conserved and is probably important in plants, especially for CG sequence contexts [28], DNA methylation at the vicinity regions of the gene body also plays a crucial role in the regulation of gene expression, such as the *cis*-regulation of genes by combining transcription factors [29]. For balanced tetrads, we found that balanced tetrads showed similar DNA methylation levels of homologous alleles from different subgenomes regardless of any sequence contexts (Figure 2D). Within non-balanced tetrads, homologous alleles with higher expression (dominant and non-suppressed) showed generally similar DNA methylation patterns to alleles with lower expression, particularly in CG sequence contexts (Figure 2E,F). However, for non-CG DNA methylation, such as in CHG and CHH contexts, homologous alleles with lower expression showed higher DNA methylation than that of alleles with higher expression at gene body regions, and these differences were not limited to the gene body regions in the CHH context, but also extended to the upstream regions. Additionally, we found that non-balanced tetrads showed higher gene body DNA methylation than that in balanced tetrads, especially for the CG context (Figure 2G). These results indicate that DNA methylation levels are correlated with homologous alleles expression bias.

3.3. Variation of TE Insertions and Selection Pressure of Tetrads

Transposable elements (TEs) are the main target sites of DNA methylation of all three sequence contexts and they could regulate the expression variation of the surrounding genes through DNA methylation. To clarify the role of TEs on DNA methylation variety between differentially expressed alleles, we calculated the abundance distribution of TEs across homologous alleles and their flanking regions. As shown in Figure 3A,B, we found significantly increased TE insertions in highly-expressed alleles in comparison to those in lowly-expressed alleles at gene body regions (p value < 0.05, Mann–Whitney test). These differences in TE insertions were also found in upstream and downstream regions. Moreover, we found that those gene body TEs were mostly inserted into intron regions, followed by exon regions. We also found that a relatively large abundance of TEs overlapped with both intron and exon regions. Because non-CG methylation was most different between non-balanced homologous alleles (Figure 2), we proposed that intronic TEs, as well as flanking TEs, could regulate host gene expression, particularly by inducing non-CG DNA methylation. This phenomenon was also observed in recent studies of rice and Arabidopsis [30,31].

To further verify our conclusion, we also examined the DNA methylation patterns between highly- (dominant and non-suppressed) and lowly-expressed (suppressed and non-dominant) alleles after excluding intronic TEs. We found no difference in non-CG methylation between highly- and lowly-expressed alleles, and the non-CG methylation levels were significantly reduced (Figure S3). We also estimated the *Ka/Ks* ratio between highly- and lowly-expressed alleles and found significant differences between them (Figure 3C). Highly-expressed alleles always showed lower *Ka/Ks* values than those in lowly-expressed alleles, and this phenomenon is consistent among eight groups of tetrads. Taken together, these findings indicate that highly-expressed alleles should undergo much stronger purifying selection pressure than lowly-expressed alleles.

3.4. Variation of CG to TG Substitution Rates among Tetrad Alleles

Many prior studies have shown that methylated cytosines are more mutable than non-methylated cytosines via deamination and bringing C:G to T:G mismatch [32,33]. To verify this association of homologous tetrad alleles with divergent expression in sugarcane, we examined the single nucleotide substitution rate by pairwise alignments of orthologous gene pairs between sugarcane and sorghum. As expected, the transitions rates were higher than transversion rates, whereas the C<=>G substitution rates were higher than other types of transversion rates. This phenomenon is consistent in balanced and non-balanced alleles (Figure 4A–E). For the transition substitutions, we found that the C<=>T substitution rate was relatively higher than the G<=>A substitution rate (Figure 4A–E). We also assessed

the C<=>T substitution rate between balanced and non-balanced alleles and found a significantly increased C<=>T substitution rate in non-balanced alleles compared to the rate in balanced alleles (Mann–Whitney test $p < 0.05$, Figure 4F). Intriguingly, we also found a very strong preference of the CG to TG substitutions compared to other 3′ nucleotide, such as CA to TA, CC to TC and CT to TT, which is consistent with the higher CG gene body DNA methylation in non-balanced alleles than that in balanced alleles (Figure 4G,H). For example, dominant, non-dominant, suppressed and non-suppressed alleles showed higher CG to TG substitution rates and increased CG gene body DNA methylation in comparison to the results observed in balanced alleles (Figures 2G and 4G,H). We further examined the relationship between genetic variations and DNA methylation changes between differentially expressed homologous alleles and found that genetic variations between homologous alleles were positively correlated with DNA methylation changes (Figure 4I). Taken together, differentially expressed homologous alleles have strong association with genetic substitution and DNA methylation.

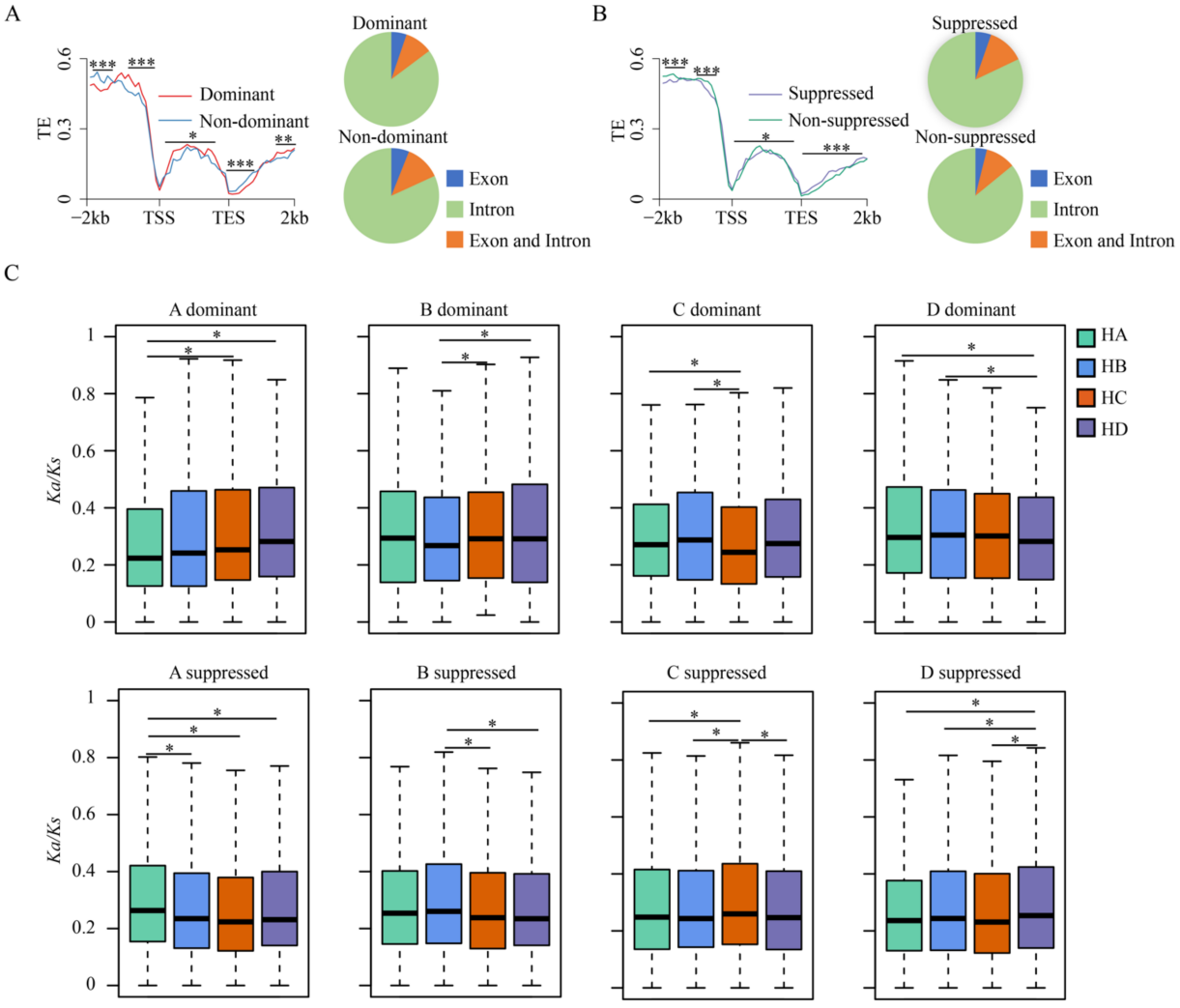

Figure 3. Transposable elements distribution and election pressure analysis. (**A**) Distribution of transposable elements between dominant and non-dominant alleles, and proportion of TEs at intron, exon or intron/exon regions. (**B**) Distribution of transposable elements between suppressed and non-suppressed alleles (**C**) *Ka/Ks* values distribution of tetrad alleles of each expression categories. The differences were tested using the Mann–Whitney test. * p value < 0.05, ** p value < 0.01, *** p value < 0.001.

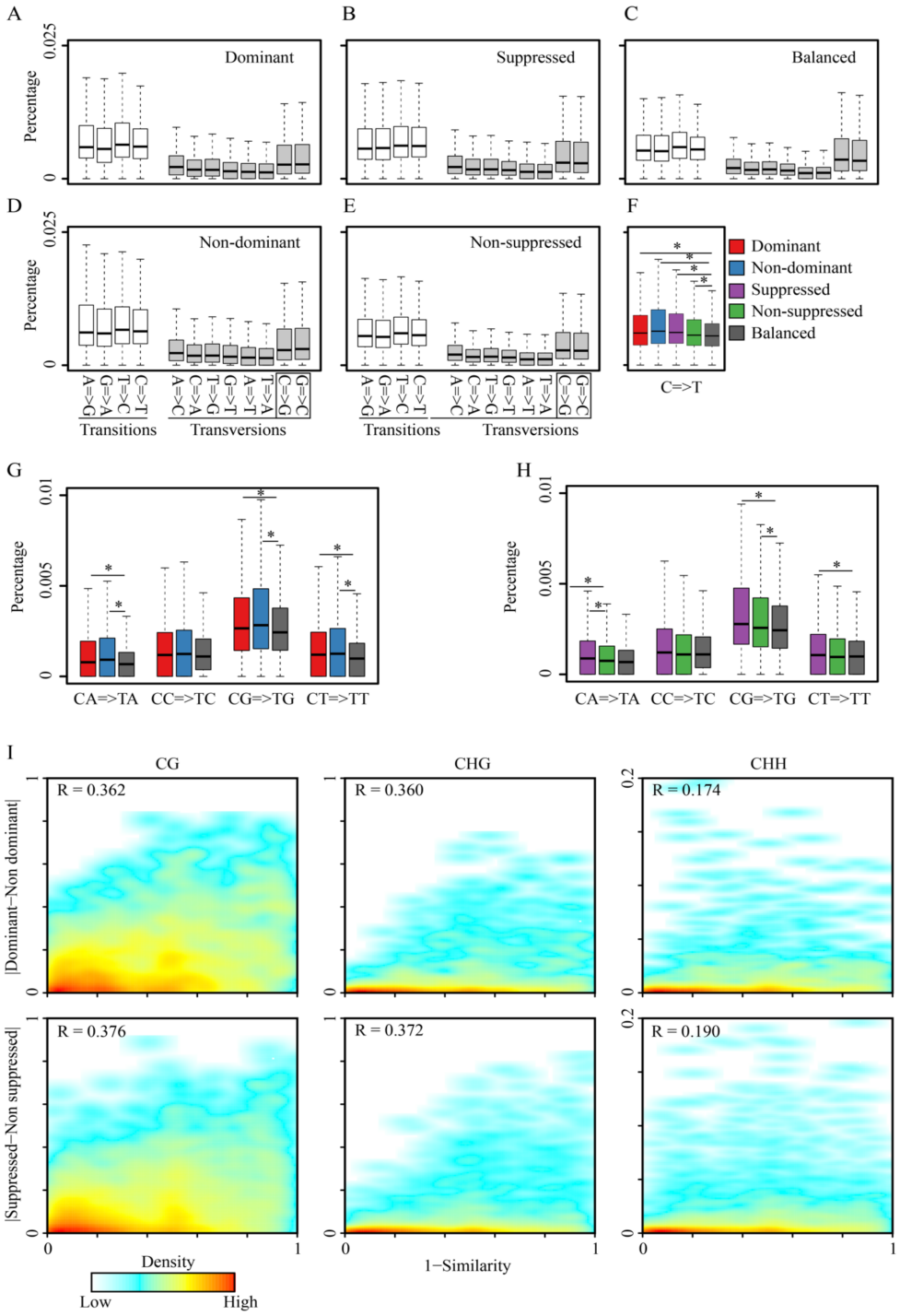

Figure 4. The mutation rate of cytosine is associated with DNA methylation. The box plot showing the mutation rate comparison of the single-nucleotide mutation of orthologous gene pairs between sugarcane and sorghum of different categories tetrads. Dominant (**A**), suppressed (**B**), balanced (**C**), non-dominant (**D**), non-suppressed (**E**). (**F**) Comparison of C<=>T substitution rate between balanced and non-balanced alleles. Non-balanced alleles were divided into dominant, non-dominant, suppressed, and non-suppressed alleles. (**G**,**H**) Comparison of the substitution rate of different mutation types between balanced and dominant/non-dominant alleles (**G**) and suppressed/non-suppressed alleles (**H**). The mutation types were separated according to cytosine 3′ neighboring nucleotide. The differences were tested using the Mann–Whitney test. * p value < 0.05. (**I**) The relationship between genetic variation and changes in DNA methylation in differentially expressed homologous alleles.

3.5. DNA Methylation Contributes to Sugarcane Agronomic Traits

DNA methylation and transcriptomic analyses between homologous alleles have revealed that transcriptional changes of homologous alleles are associated with DNA methylation changes (Figure 2). To investigate the contribution of DNA methylation changes to genes involved in important agronomic traits of sugarcane, we examined the differentially expressed methylation-associated alleles, and found that a large number of differentially expressed alleles were associated with changes in DNA methylation, either of sequence contexts (CG, CHG, and CHH) or location (upstream, gene body, and downstream regions). CG and CHG methylation play more important roles in regulation alleles than CHH methylation either dominant or suppressed alleles (Figure 5A,B). In addition, most of the alleles were not regulated by all of the sequence contexts of DNA methylation. In other words, only a small proportion of alleles were regulated by all three sequence DNA methylation contexts, such as 194 alleles out of 2160 and 620 alleles out of 6570 in the dominant and suppressed groups, respectively (Figure 5C,D). This indicates that allele expression is regulated by context-specific DNA methylation.

Expression changes in homologous alleles are expected to affect protein expression, and ultimately phenotype [14,34]. Early analysis of duplicated genes in cassava showed that differentially expressed duplicated genes regulated by changes of DNA methylation were always enriched into important agronomic pathways, such as carbohydrate metabolism [35]. We performed the functional enrichment analysis of non-balanced alleles regulated by DNA methylation and found that several agronomic traits related pathways were enriched in dominant and suppressed group alleles. In dominant group alleles, peptide metabolic processes, amino acid activation and several responses to stimuli were enriched (Figure 5E, and Table S6), but in suppressed group alleles, protein metabolism-related pathways were enriched, such as protein localization, protein transport, and protein complex assembly (Figure 5F, and Table S7). For example, Sspon.03G0004720 tetrad (−1A, −2B, −3C, and −4D four alleles), which was involved in heat acclimation, protein SUMOylation, and response to heat, was classified into the HD dominant group across all four tissues (Figure 5G). We found that these four alleles showed differential DNA methylation at upstream and downstream regions, and we proposed that this different DNA methylation might explain the expression divergence of these alleles. Collectively, the association between DNA methylation and the transcription divergence of homologous alleles might affect many alleles involved in the pathways related to the agronomic traits of sugarcane during domestication.

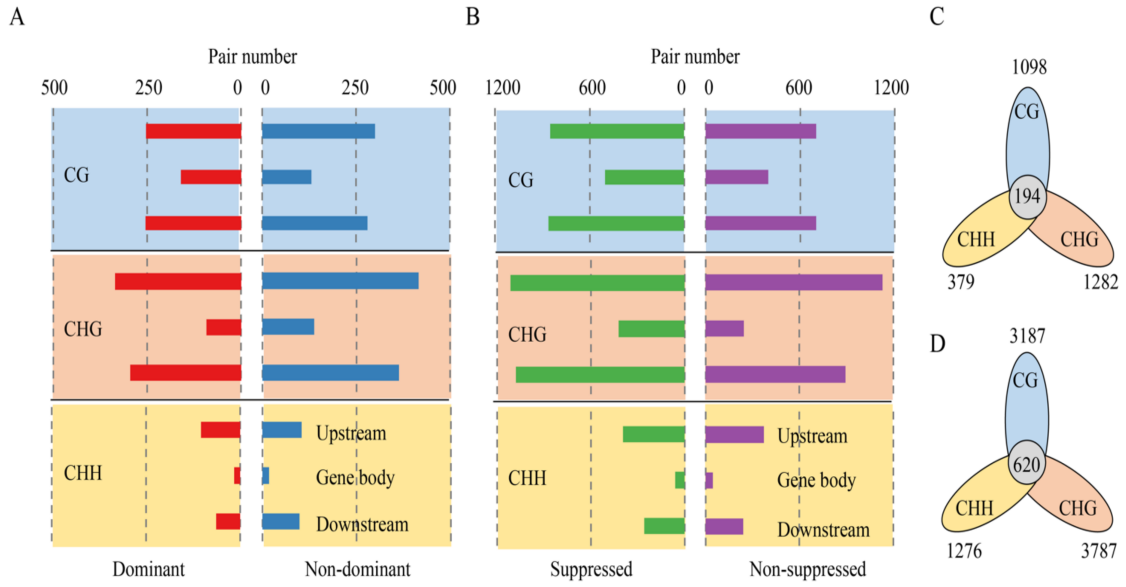

Figure 5. *Cont.*

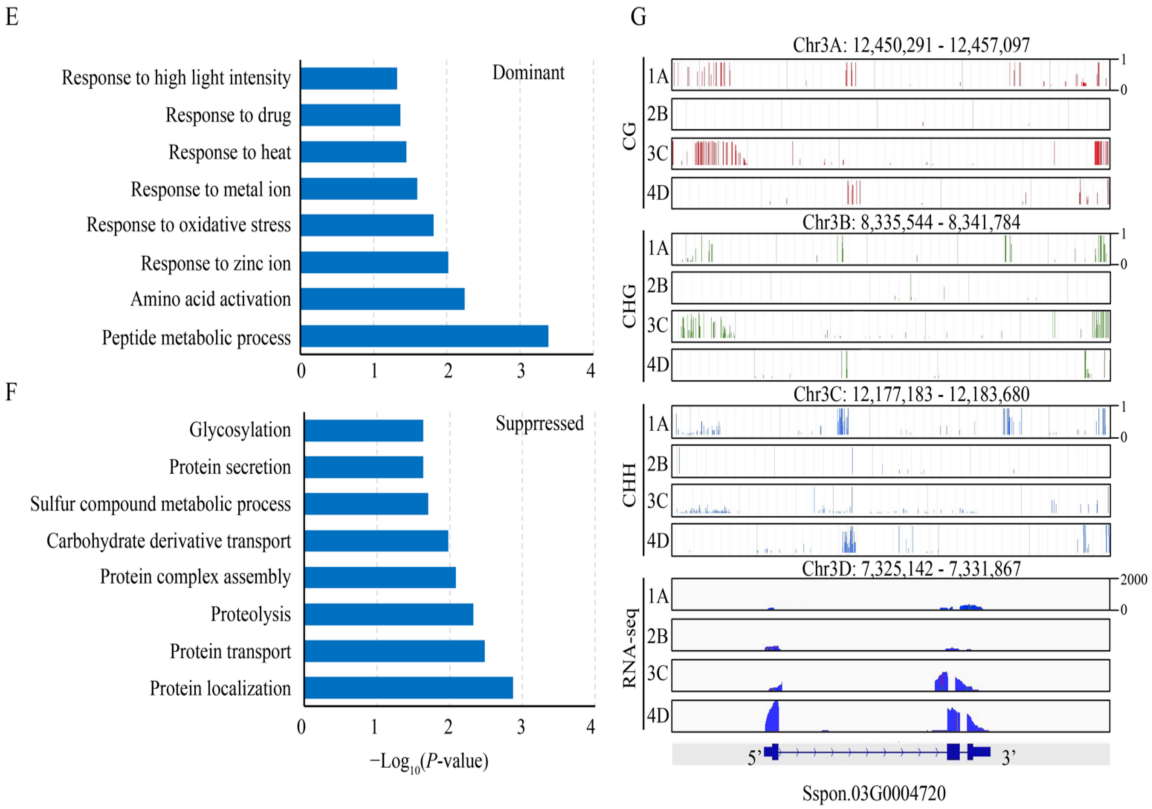

Figure 5. Allele-specific expression and DNA methylation. (**A**) DNA methylation divergence between dominant and non-dominant tetrad alleles in CG, CHG, and CHH sequence contexts. (**B**) DNA methylation divergence between suppressed and non-suppressed tetrad alleles in CG, CHG, and CHH sequence contexts. (**C**,**D**) Venn diagram showing the overlap of alleles regulated by DNA methylation in all three contexts from dominant categories (**C**) and suppressed categories (**D**). (**E**) Enriched GO terms of DNA methylation-associated dominant category alleles. (**F**) Enriched GO terms of DNA methylation-associated suppressed category alleles. (**G**) Snapshot of example dominant tetrad alleles with varied DNA methylation.

4. Discussions and Conclusions

DNA methylation is a stable epigenetic modification that can affect gene expression and transposable element activities, and it plays an important role in many biological progresses, such as development, stress response, and genome evolution [36–39]. The dynamic DNA methylation patterns during subgenome evolution and homoeologs expression regulation has been demonstrated to be a critical regulation mechanism in polyploid plants [13–15]. Differential expression homoeologs were always associated with changes in DNA methylation, especially for allopolyploids, such as *B. napus*, wheat, and cotton [13,16,17]. However, there is still a lack of comprehensive DNA methylation studies in the evolution of the genomes of autopolyploid plants. *S. spontaneum* is an autopolyploid species which was assembled recently and annotated into four groups of homologous chromosomes with homologous alleles defined [2]. In this study, we have generated and examined DNA methylation and transcriptome data across four tissues of sugarcane.

Our data clearly showed that there is no genome-wide expression difference between homologous alleles among different homologous chromosomes, as well as DNA methylation. Tetrads are the specific characteristic of haploid *S. spontaneum*, which is an octoploid, so we mainly focused on tetrads in the following study. In particularly, the majority of tetrad alleles showed non-balanced expression, which is consistent across the four different tissues. In contrast to other polyploid studies, most of the homoeologs are in balanced expression between the subgenomes [12,22,40]. For example, in wheat, ~30% of homoeolog triads showed non-balanced expression patterns, with higher or lower expression from a

single homoeolog with respect to the other two [22]. This might indicate species-specific differences or auto-/allo-polyploids differences. To further investigate the association between homologous allele expression changes and DNA methylation divergence, we focused on tetrad alleles and classified them into nine categories similar to the study of allopolyploid wheat [22]. Balanced and non-balanced tetrad alleles showed very different expression patterns among homologous chromosomes, and a high proportion of non-balanced alleles were found in the present study, which is different to the studies in other allopolyploids. In wheat and cotton, approximately 30% of the homoeologs were classified into the non-balanced group, but more than 80% of the tetrad alleles were classified into the non-balanced group in sugarcane. This may be partly explained by only comparing the tetrad alleles, because we found that the majority of the homologous alleles were not expressed differently by pairwise comparison, (Figures 1B and S1B–F). Gene expression was always associated with DNA methylation. We found that the non-balanced alleles showed significantly increased DNA methylation in comparison to the balanced alleles, particularly for non-CG methylation. However, we did not find a significant DNA methylation difference between the tetrad alleles even though there are significant expression changes between them, indicating that additional factors may be involved in the process [41]. Recent studies indicate that histone modification and small RNAs have diverse roles in homoeologs expression changes in allopolyploids [13,42]. Small RNAs may have important roles in allotetraploid formation, as changes in the expression or accumulation of small RNAs are observed in inter-specific hybrids [43,44]. Further progress in DNA methylation, histone modification and small RNAs in regard to the regulation of duplicates' expression in polyploids will help us to understand the genome expression and epigenetic regulation in polyploid genomes and will provide epigenetic insights into manipulating agronomical traits in polyploid crops.

Recent studies have shown that the expression divergence of homologous genes is associated with genetic variation [45]. Consistent with these findings, the expression difference between homologous alleles in sugarcane is correlated with genetic variation at both gene body and vicinity regions. Meanwhile, we observed that transposons inserted into intron or upstream regions might play a crucial role in this correlation. This is consistent with recent studies showing that intronic TEs can cause changes of host gene expression by altering epigenetic state [31]. In addition, highly-expressed alleles always showed smaller *Ka/Ks* values than those of lowly-expressed alleles and they tend to be under strong purifying selection pressure. Although DNA methylation is important for gene expression regulation, it also brings about some adverse effects in the sense that methylated cytosines are more prone to spontaneous deamination than unmethylated cytosines and deamination to thymine [32,33]. Therefore, methylated cytosines bear a high rate of mutation, especially cytosine to thymine transitions. In the comparison between balanced and non-balanced alleles, variation of CG to TG substitution rates were observed, this might reflect the differences in DNA methylation between balanced and non-balanced alleles, particularly in the CG context. This comparison demonstrated the complicated association between DNA methylation, expression, transposons insertion, and mutation rate.

Growing evidence has suggested that DNA methylation plays a crucial role in the regulation of development and phenotypic diversity in plants. In a recent study, DNA methylation variation in CONSTANS-like (COL) genes was found to be responsible for the loss of photoperiod sensitivity during cotton domestication [14]. The loss of DNA methylation of a LINE retrotransposon (*Karma*) in the intron of the homeotic gene *DEFICIENS* causes alternative splicing and premature termination and consequently the generation of the unproductive mantled somaclonal variant of oil palm [34]. In the present study, our findings showed that many homologous alleles regulated by changes in DNA methylation are involved in several agronomic trait-related pathways, such as the response to heat stress. These studies suggest that DNA methylation variation is an important component of artificial selection in crop domestication beyond genetic variation.

Taken together, our results demonstrate that epigenetic regulation, such as DNA methylation, is likely to be involved in the expression divergence of plenty of homologous alleles, which are involved in several important biological processes, and therefore has potential value for improving agronomic performance.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy12112852/s1, Figure S1: Pairwise expression comparison between homologous alleles of different subgenomes; Figure S2: Association between DNA methylation and gene expression of homologous alleles; Figure S3: DNA methylation variation and association with expression of tetrad alleles; Table S1: Summary of nine expression categories of homologous tetrad alleles in four tissues; Table S2: GO enrichment analysis of balanced tetrad alleles; Table S3: GO enrichment analysis of dominant tetrad alleles; Table S4: GO enrichment analysis of suppressed tetrad alleles; Table S5: GO enrichment analysis of different DNA methylation association to dominant tetrad alleles; Table S6: GO enrichment analysis of different DNA methylation association to dominant tetrad alleles; Table S7: GO enrichment analysis of different DNA methylation association to suppressed tetrad alleles.

Author Contributions: H.W. and B.C. designed and coordinated the study. Y.X., Y.S. and Y.Q. performed methylome and transcriptome data analysis. C.Z., H.Y., and C.S. collected tissues and coordinated the sequencing. Y.X. and X.C. made contributions to the experimental validation. H.W. and Y.X. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 32160142) and the Sugarcane Research Foundation of Guangxi University (Grant No. 2022GZA002) to H.W. B.C was supported by grant from the Department of Science and Technology of Guangxi Zhuang Autonomous Region (AD17129002). Y.X. and H.Y were supported by the Innovation Project of Guangxi Graduate Education (YCBZ2021005).

Data Availability Statement: Sequence data are deposited in the Sequence Read Archive (SRA) database with the study accession number PRJNA730638.

Acknowledgments: We thank Jiangfeng Huang, Xiping Yang at College of Agriculture, Guangxi University for providing plant materials, Liang Xie for helpful discussion with DNA methylation analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meena, M.R.; Appunu, C.; Arun Kumar, R.; Manimekalai, R.; Vasantha, S.; Krishnappa, G.; Kumar, R.; Pandey, S.K.; Hemaprabha, G. Recent Advances in Sugarcane Genomics, Physiology, and Phenomics for Superior Agronomic Traits. *Front. Genet.* **2022**, *13*, 854936. [CrossRef] [PubMed]
2. Zhang, J.; Zhang, X.; Tang, H.; Zhang, Q.; Hua, X.; Ma, X.; Zhu, F.; Jones, T.; Zhu, X.; Bowers, J.; et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **2018**, *50*, 1565–1573. [CrossRef] [PubMed]
3. Garsmeur, O.; Droc, G.; Antonise, R.; Grimwood, J.; Potier, B.; Aitken, K.; Jenkins, J.; Martin, G.; Charron, C.; Hervouet, C.; et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* **2018**, *9*, 2638. [CrossRef] [PubMed]
4. Zhang, L.; Chen, F.; Zhang, X.; Li, Z.; Zhao, Y.; Lohaus, R.; Chang, X.; Dong, W.; Ho, S.Y.W.; Liu, X.; et al. The water lily genome and the early evolution of flowering plants. *Nature* **2020**, *577*, 79–84. [CrossRef] [PubMed]
5. Edger, P.P.; Poorten, T.J.; VanBuren, R.; Hardigan, M.A.; Colle, M.; McKain, M.R.; Smith, R.D.; Teresi, S.J.; Nelson, A.D.L.; Wai, C.M.; et al. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **2019**, *51*, 541–547. [CrossRef]
6. Zhuang, W.; Chen, H.; Yang, M.; Wang, J.; Pandey, M.K.; Zhang, C.; Chang, W.C.; Zhang, L.; Zhang, X.; Tang, R.; et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* **2019**, *51*, 865–876. [CrossRef]
7. Otto, S.P. The evolutionary consequences of polyploidy. *Cell* **2007**, *131*, 452–462. [CrossRef]
8. One Thousand Plant Transcriptomes, I. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **2019**, *574*, 679–685. [CrossRef]
9. Ren, R.; Wang, H.; Guo, C.; Zhang, N.; Zeng, L.; Chen, Y.; Ma, H.; Qi, J. Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. *Mol. Plant* **2018**, *11*, 414–428. [CrossRef]

10. Wang, H.; Guo, C.; Ma, H.; Qi, J. Reply to Zwaenepoel et al.: Meeting the Challenges of Detecting Polyploidy Events from Transcriptomic Data. *Mol. Plant* **2019**, *12*, 137–140. [CrossRef]
11. Yoo, M.J.; Szadkowski, E.; Wendel, J.F. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **2013**, *110*, 171–180. [CrossRef]
12. Pfeifer, M.; Kugler, K.G.; Sandve, S.R.; Zhan, B.; Rudi, H.; Hvidsten, T.R.; International Wheat Genome Sequencing, C.; Mayer, K.F.; Olsen, O.A. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **2014**, *345*, 1250091. [CrossRef]
13. Song, Q.; Guan, X.; Chen, Z.J. Dynamic Roles for Small RNAs and DNA Methylation during Ovule and Fiber Development in Allotetraploid Cotton. *PLoS Genet.* **2015**, *11*, e1005724. [CrossRef]
14. Song, Q.; Zhang, T.; Stelly, D.M.; Chen, Z.J. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* **2017**, *18*, 99. [CrossRef]
15. Cheng, F.; Sun, C.; Wu, J.; Schnable, J.; Woodhouse, M.R.; Liang, J.; Cai, C.; Freeling, M.; Wang, X. Epigenetic regulation of subgenome dominance following whole genome triplication in Brassica rapa. *New Phytol.* **2016**, *211*, 288–299. [CrossRef]
16. Li, Z.; Wang, M.; Lin, K.; Xie, Y.; Guo, J.; Ye, L.; Zhuang, Y.; Teng, W.; Ran, X.; Tong, Y.; et al. The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol.* **2019**, *20*, 139. [CrossRef]
17. Yuan, Y.; Yang, X.; Feng, M.; Ding, H.; Khan, M.T.; Zhang, J.; Zhang, M. Genome-wide analysis of R2R3-MYB transcription factors family in the autopolyploid Saccharum spontaneum: An exploration of dominance expression and stress response. *BMC Genom.* **2021**, *22*, 622. [CrossRef]
18. Wang, L.; Xie, J.; Hu, J.; Lan, B.; You, C.; Li, F.; Wang, Z.; Wang, H. Comparative epigenomics reveals evolution of duplicated genes in potato and tomato. *Plant J.* **2018**, *93*, 460–471. [CrossRef]
19. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
20. Xi, Y.; Li, W. BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinform.* **2009**, *10*, 232. [CrossRef]
21. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]
22. Ramirez-Gonzalez, R.H.; Borrill, P.; Lang, D.; Harrington, S.A.; Brinton, J.; Venturini, L.; Davey, M.; Jacobs, J.; van Ex, F.; Pasha, A.; et al. The transcriptional landscape of polyploid wheat. *Science* **2018**, *361*, eaar6089. [CrossRef]
23. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Zhang, J.; Saski, C.A.; Scheffler, B.E.; Stelly, D.M.; et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [CrossRef]
24. Chalhoub, B.; Denoeud, F.; Liu, S.; Parkin, I.A.; Tang, H.; Wang, X.; Chiquet, J.; Belcram, H.; Tong, C.; Samans, B.; et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **2014**, *345*, 950–953. [CrossRef]
25. Yang, J.; Liu, D.; Wang, X.; Ji, C.; Cheng, F.; Liu, B.; Hu, Z.; Chen, S.; Pental, D.; Ju, Y.; et al. The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* **2016**, *48*, 1225–1232. [CrossRef]
26. International Wheat Genome Sequencing, C. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **2014**, *345*, 1251788. [CrossRef]
27. Wang, L.; Shi, Y.; Chang, X.; Jing, S.; Zhang, Q.; You, C.; Yuan, H.; Wang, H. DNA methylome analysis provides evidence that the expansion of the tea genome is linked to TE bursts. *Plant Biotechnol. J.* **2019**, *17*, 826–835. [CrossRef] [PubMed]
28. Niederhuth, C.E.; Bewick, A.J.; Ji, L.; Alabady, M.S.; Kim, K.D.; Li, Q.; Rohr, N.A.; Rambani, A.; Burke, J.M.; Udall, J.A.; et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* **2016**, *17*, 194. [CrossRef] [PubMed]
29. Crisp, P.A.; Marand, A.P.; Noshay, J.M.; Zhou, P.; Lu, Z.; Schmitz, R.J.; Springer, N.M. Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 23991–24000. [CrossRef] [PubMed]
30. Espinas, N.A.; Tu, L.N.; Furci, L.; Shimajiri, Y.; Harukawa, Y.; Miura, S.; Takuno, S.; Saze, H. Transcriptional regulation of genes bearing intronic heterochromatin in the rice genome. *PLoS Genet.* **2020**, *16*, e1008637. [CrossRef]
31. Le, T.N.; Miyazaki, Y.; Takuno, S.; Saze, H. Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res.* **2015**, *43*, 3911–3921. [CrossRef]
32. Ossowski, S.; Schneeberger, K.; Lucas-Lledo, J.I.; Warthmann, N.; Clark, R.M.; Shaw, R.G.; Weigel, D.; Lynch, M. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **2010**, *327*, 92–94. [CrossRef]
33. Kiefer, C.; Willing, E.M.; Jiao, W.B.; Sun, H.; Piednoel, M.; Humann, U.; Hartwig, B.; Koch, M.A.; Schneeberger, K. Interspecies association mapping links reduced CG to TG substitution rates to the loss of gene-body methylation. *Nat. Plants* **2019**, *5*, 846–855. [CrossRef]
34. Ong-Abdullah, M.; Ordway, J.M.; Jiang, N.; Ooi, S.E.; Kok, S.Y.; Sarpan, N.; Azimi, N.; Hashim, A.T.; Ishak, Z.; Rosli, S.K.; et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **2015**, *525*, 533–537. [CrossRef]
35. Wang, H.; Beyene, G.; Zhai, J.; Feng, S.; Fahlgren, N.; Taylor, N.J.; Bart, R.; Carrington, J.C.; Jacobsen, S.E.; Ausin, I. CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 13729–13734. [CrossRef]

36. Dowen, R.H.; Pelizzola, M.; Schmitz, R.J.; Lister, R.; Dowen, J.M.; Nery, J.R.; Dixon, J.E.; Ecker, J.R. Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2183–E2191. [CrossRef]
37. Quadrana, L.; Almeida, J.; Asis, R.; Duffy, T.; Dominguez, P.G.; Bermudez, L.; Conti, G.; Correa da Silva, J.V.; Peralta, I.E.; Colot, V.; et al. Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* **2014**, *5*, 3027. [CrossRef]
38. Secco, D.; Wang, C.; Shou, H.; Schultz, M.D.; Chiarenza, S.; Nussaume, L.; Ecker, J.R.; Whelan, J.; Lister, R. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *Elife* **2015**, *4*, e09343. [CrossRef]
39. Klopfenstein, D.V.; Zhang, L.; Pedersen, B.S.; Ramírez, F.; Warwick Vesztrocy, A.; Naldi, A.; Mungall, C.J.; Yunes, J.M.; Botvinnik, O.; Weigel, M.; et al. GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **2018**, *8*, 10872. [CrossRef]
40. Wu, J.; Lin, L.; Xu, M.; Chen, P.; Liu, D.; Sun, Q.; Ran, L.; Wang, Y. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genom.* **2018**, *19*, 586. [CrossRef]
41. Chodavarapu, R.K.; Feng, S.; Ding, B.; Simon, S.A.; Lopez, D.; Jia, Y.; Wang, G.L.; Meyers, B.C.; Jacobsen, S.E.; Pellegrini, M. Transcriptome and methylome interactions in rice hybrids. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12040–12045. [CrossRef] [PubMed]
42. Song, Q.; Chen, Z.J. Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* **2015**, *24*, 101–109. [CrossRef] [PubMed]
43. Jiao, W.; Yuan, J.; Jiang, S.; Liu, Y.; Wang, L.; Liu, M.; Zheng, D.; Ye, W.; Wang, X.; Chen, Z.J. Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. *Plant J.* **2018**, *93*, 828–842. [CrossRef] [PubMed]
44. Li, A.; Liu, D.; Wu, J.; Zhao, X.; Hao, M.; Geng, S.; Yan, J.; Jiang, X.; Zhang, L.; Wu, J.; et al. mRNA and Small RNA Transcriptomes Reveal Insights into Dynamic Homoeolog Regulation of Allopolyploid Heterosis in Nascent Hexaploid Wheat. *Plant Cell* **2014**, *26*, 1878–1900. [CrossRef]
45. Hu, W.; Ji, C.; Shi, H.; Liang, Z.; Ding, Z.; Ye, J.; Ou, W.; Zhou, G.; Tie, W.; Yan, Y.; et al. Allele-defined Genome Reveals Biallelic Differentiation during Cassava Evolution. *Mol. Plant* **2021**, *14*, 851–854. [CrossRef]