DNA methylome and lncRNAome Analysis Provide Insights Into Mechanism of Genome-dosage Effects in Autotetraploid Cassava

Liang Xiao

Guangxi Academy of Agricultural Science

Liuying Lu

Guangxi Academy of Agricultural Science

Wendan Zeng

Guangxi Academy of Agricultural Science

Xiaohong Shang

Guangxi Academy of Agricultural Science

Sheng Cao

Guangxi Academy of Agricultural Science

Huabing Yan (✉ h.b.yan@hotmail.com)

Guangxi Academy of Agricultural Science <https://orcid.org/0000-0002-2991-3597>

Research

Keywords: Autotetraploid cassava, Genome-dosage effect, DNA methylation, lncRNA expression, Protein coding gene expression, Transposon

Posted Date: October 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-934244/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

During newly formed polyploidy, one of the most intriguing aspects is that whole-genome duplication (WGD) increase the dosage of all coding and non-coding genes. However, the molecular implications of genome-dosage effects remain elusive.

Results

We conducted integrated maps of methylomes and lncRNAomes in autotetraploid cassava (*Manihot esculenta* Crantz) and its donor parent, both of which were independently clonal propagated for three years. DNA methylation variation of transposable elements (TEs) was observed as widespread in autotetraploid cassava. The hypermethylation of DNA transposons in mCG and mCHH sites may be an effective way to suppress the expression of nearby PCGs in autotetraploid cassava, resulting in similar expression levels for most of PCGs between autotetraploid and diploid cassava. The decreased methylation levels of retrotransposons in mCHG and mCHH sites, which partly attributed to reduction methylation of *Cypsy* neighboring long intergenic noncoding RNAs in autotetraploid cassava, may be a mechanism that may suppress the expression levels of nearby lncRNA, leading to no significant differences in transcriptome alterations for major of lncRNAs from its diploid parent.

Conclusions

This work highlighted that WGD-induced DNA methylation variation in DNA transposons and retrotransposons may be as direct adaptive responses to dosage of all coding-genes and lncRNAs, respectively.

Introduction

Polyploidy or whole-genome duplication (WGD), is the heritable condition of possessing multiple sets of chromosomes co-occur in a nucleus, with four being the most common [1, 2]. Polyploidy has long been recognized as a major force in angiosperm evolution and can result in considerable changes in both coding and non-coding genes expression, which provided a molecular basis for well adapted [3-6]. Two forms of polyploidy are often considered in plants [7]: allopolyploidy species are traditionally considered to arise via interspecific hybridization and subsequent doubling of non-homologous genomes (AABB) [8, 9], whereas autopolyploids arise within a single species by doubling of structurally similar, homologous genomes (AAAA) [10, 11]. Although past views pointed that autopolyploidy is likely rare, increasing evidences indicated that autopolyploid taxa might be more common and the appearance of autotetraploidy plants in nature might be significantly underestimated [12-16], despite potential weaknesses, such as sterility, aneuploidy, genomic and epigenetic instabilities [1, 2]. During newly formed polyploids, one of the most intriguing aspects is that genome doubling events increase the dosage of all coding and non-coding genes. However, the putative mechanism for this process is largely

unknown. Autopolyploids, especially artificial lines, ruling out disturbances from incompatible genomes, offer an extraordinary opportunity to understand mechanism of genome-dosage effects.

Synthesized autopolyploid without many changes in DNA sequence were subjected to a little functional reorganization of gene expression during the first generations after genome double *per se* [15]. Transcriptome analyses have shown that subtle changes in regulatory networks, only negligible differences of protein coding genes (PCGs) and long non coding RNA (lncRNA) expression are found between synthesized autopolyploids and their progenitors [17-24]. However, what control the changes of PCGs and lncRNAs expression contributing to dominant phenotype in autotetraploids is poorly understood.

Doubling a set of chromosome cause 'genome shock', associated with dramatic changes in the epigenetic modifications [25, 26]. DNA methylation provides an effective mean for a polyploid cell to overcome 'genomic shock' caused by WGD [27, 28]. Cytosine methylation is a common feature of epigenetic regulation that influences many molecular processes, including embryogenesis [29], transposable elements (TEs) activity [30-32] and gene expression [33]. Plant genomes is often methylated in CG, CHG, and CHH (H = A, T, or C) contexts [34]. Many studies in allopolyploid crop species indicated that gene expression is altered more by interspecific hybridization than by polyploidization [8-9, 35-36]. However, to date, there are almost no reports on DNA methylation variations to reveal the impact on PCGs expression responding to autotetraploidization except in rice (*Oryza sativa* ssp. *Indica*) [21].

The complement of TEs within any one genome typically includes both class I retrotransposon and class II DNA transposons [37]. The latest studies have reported that TEs make up a substantial fraction of mature lncRNA transcripts, they are also enriched in the vicinity of lncRNAs, where they frequently contribute to their transcriptional regulation [38-40]. Zhao et al. [41] reported that demethylated *LINEs*/TEs might distinctively impact the lncRNAs expression in polyploid cotton interspecific F₁ hybrid in genomic shock caused by interspecific hybrid and WGD. Nevertheless, the impact on the lncRNAs expression of whole genome caused by WGD, especially TE-overlapped lncRNAs, remains largely unknown.

Cassava (*Manihot esculenta* Crantz), a perennial shrub of the Euphorbiaceae family, is ranked the third most consumed carbohydrate source for millions of people in the tropics [42, 43]. Cassava can be used as a good source of biofuels and alternative energy with low-planting costs. The advantages of cassava have made it a great prospect for industrial applications. The cassava genome is highly heterozygous because of its outcrossing nature and broad tropical distribution [44-46]. Previously, we obtained an autotetraploid cassava line from the diploid cultivar 'Xinxuan 048', which was produced by colchicine-induced [47]. The synthesized autopolyploid consistently exhibits significant phenotypic changes, compared to its diploid progenitor. The transcriptome data results indicated that only percentage of three point one differential expressed PCGs and one point one percent differential expressed lncRNAs were observed between autotetraploid and diploid cassava [23]. Synthesized autotetraploid, together with

diploid parent, is a fine genetic model to provide insights into genomic changes that occur in response to ploidy level. Here, we conducted integrated maps of methylomes and lncRNAomes in autotetraploid cassava and its donor parent, both of which were independently clonal propagated for three years, in order to evaluate the short-term impact of intraspecies genome duplication on the expression of PCGs and lncRNAs of whole genomes.

Results

Distinct phenotypes of autotetraploid cassava compared with its diploid progenitor

Autopolyploid cassava (4x) had previously been generated by colchicine-induced chromosome doubling from the diploid (2x) cultivar 'Xinxuan 048' [47]. The ploidy levels of the generated 4x plants were detected by the flow cytometry analysis, and then chromosome counting in root-tip cells confirmed the results of flow cytometry analysis. The 2x cassava had 36 chromosomes and the 4x had 72 chromosomes [47]. The plants were propagated and maintained for two years. The leaves of the 4x cassava were significantly larger, greater and darker green than those of 2x cassava (Additional file 1: Fig. S1) and showed improved drought resistance [23].

Single base-resolution maps of DNA methylation for diploid and autotetraploid cassava

In order to investigate the potential role of DNA methylation in response to autotetraploidy, the methylomes of 2x and 4x cassava leaf were decoded and analyzed. To assess variability, three biological replicates were generated of 2x and 4x, respectively. Genome mapping analysis showed that 113,078,644 (73.50%), 109,732,951 (73.44%), and 111,434,206 (74.25%) clean reads of each three replicate of 2x sample, while 119,723,975 (74.70%), 96,864,854 (75.06%) and 110,185,024 (75.65%) clean reads could be mapped for each replicate library. The sequence depth of all the samples were more than 24x. More than 99% of cytosines were altered, which indicated that a high rate of conversion (Additional file 2: Table S1). In addition, pearson correlation coefficients between the three replicates of 2x and 4x were found to be between 0.96 and 0.97 (Additional file 3: Fig. S2). All the data indicated that the quality of sequencing was satisfactory with subsequent analysis.

Among all the sequenced C sites, the 2x and 4x genome presented 67.09% and 66.94% (mCGs), 49.03% and 48.67% (mCHG), and 5.88% and 5.74% (mCHH) in three sequences contexts, respectively (Additional file 4: Fig. S3), which reflected all three contexts methylation did not show any significant variation after WGD ($P > 0.05$). At the chromosome-scale scale, it was discovered that the methylation levels of three contexts were all predominantly highly pericentromeric heterochromatin regions and methylation levels of all three contexts in 2x and 4x cassava were similar to one another (Additional file 5: Fig. S4).

Landscape of PCGs, lncRNAs and TEs methylation

To characterize the DNA methylation patterns in different cassava genomic regions, we constructed the methylation profiles within PCGs, lncRNAs and TEs, together with 2-kb regions flanking the genes, using

the same cut-off lengths (Fig. 1). For PCGs, lncRNAs and TE regions, relatively high methylation levels were identified in CG context, followed by CHG and CHH contexts. Methylation patterns across PCGs (Fig. 1a) and TEs (Fig. 1c) in our study was consistent with that of Wang et al. [48]. The average methylation levels of TEs were much higher than that of PCGs and lncRNAs, indicating TEs were easier to be methylated. There were no differences of CG methylation levels in PCGs bodies in 2x and 4x cassava (Wilcoxon rank sum test, P value = 0.7465, n = 33,033) (Fig. 1a). For mCHG and mCHH sites, hypermethylation state of PCGs bodies was observed in 2x cassava (Wilcoxon rank sum test, mCHG P value = 0.00010365, mCHH P value = 9.76E-45, n = 33,033) (Fig. 1a). The methylation state of lncRNA in all three contexts were higher than those of PCGs (Fig. 1b). Similarity, 4x cassava had decreased CG, CHG and CHH methylation levels relative to 2x across lncRNA body regions and flanking regions (Wilcoxon rank sum test, mCG P value = 0.0269798, mCHG P value = 0.73398116, mCHH P value = 0.0118822, n = 13,533) (Fig. 1b). We also found that TEs had decreased methylation levels in CHG and CHH contexts in 4x cassava compared to 2x (Wilcoxon rank sum test, mCG P value = 8.98E-06, mCHG P value = 4.23E-21, mCHH P value = 5.13E-75, n = 617,861) (Fig. 1c). In summary, WGD may have widespread influence on methylation levels of both PCGs and lncRNAs body regions in CHG and CHH contexts, rather than CG context.

Consequently, we detected the methylation differences of the two classes of TEs. All the 13 orders of TEs had unique average methylation distribution, and exhibited hypermethylation state in body regions than flanking regions in all three contexts. In light of the proportion of *SINE*, *Stowaway*, *Harbinger* and *other_DNA* is too small in the cassava genome, so the four types of TEs were not be considered for further analysis (Additional file 6: Table S2). Obviously, all the class I TEs including *Copia*, *Gypsy*, *LINE*, and *other_LTR* exhibited hypermethylation levels in mCHG and mCHH sites in 2x cassava (Fig. 1d, e). We also found that almost all the body regions of class II TEs from 4x cassava had increased CG methylation levels except those of *En_Spm* and *MuLE_MuDR*. Moreover, body regions of *MITE* from 4x cassava were hypermethylated in all three contexts. Although *En_Spm*, *Helitron*, *hAT* and *MULE-MuDR* exhibited hypermethylation levels in mCHG and mCHH sites in 2x cassava, methylation levels in CHG and CHH contexts were much lower than those in CG context (Fig. 1f, g).

Unlike class I TEs, class II TEs tended to localize in euchromatin regions where PCGs were actively expressed. Genome-wide changes of TEs methylation levels may affect expression of neighboring PCGs that were inserted and surrounded by class II TEs after WGD. On the other hand, the proportion of TE-overlapped lncRNAs to all lncRNAs detected is 40% (Additional file 7: Fig. S5), which supported a major of lncRNAs are either derived from TEs or contain TEs remnants [38-40]. Genome-wide alteration of class I TEs methylation levels may affect expression of adjacent lncRNAs as a result of autotetraploidization. Therefore, it was sensible and necessary to combine TE methylation and PCG and lncRNA expression to examine the epigenetic responses to WGD.

Gene methylation is associated with gene activity

In view of the difference of methylation between 2x and 4x cassava after genome doubling, we attempted to understand whether PCG- and lncRNA-expression levels were influenced by DNA methylation. A total of 33,030 PCGs and 13,517 lncRNAs assembled from the lncRNA-seq data in 'Xinxuan 048' were divided into four quartiles from high-expressed group, low-expressed group, middle-expressed group and non-expressed group based on their expression levels according to the criteria of Yan et al. [49]. PCGs with relative high CG body methylation level showed high expression, but relatively low expression in the flanking regions in 2x and 4x cassava (Fig. 2a). In contrast, PCGs with high-expressed showed the lowest CHG and CHH methylation levels, middle-expressed PCGs displayed the second highest CHG methylation levels, and PCGs with non-expressed displayed the highest methylation level in the two cassava genotypes (Fig. 2b, c). The results indicated that positive correlation was observed between DNA CG methylation and PCG-expression levels, there was negative correlation for CHG and CHH methylation levels and PCG expression in 2x and 4x cassava.

As is shown in Fig. 2d, there was irregular of DNA methylation of mCG against global lncRNA expression because the four expressional graph against the mCG site of lncRNAs twined together seriously in 2x and 4x cassava. lncRNAs with high-expressed showed the highest CHG methylation levels, middle-expressed lncRNAs displayed the second highest CHG methylation levels, and lncRNAs with non-expressed displayed almost the same relative low methylation level as that of the low-expressed lncRNAs. CHG methylation throughout lncRNA body regions together with flanking regions is positive with lncRNA expression. mCHH sites of upstream and downstream regions is positive correlated with lncRNA expression, lncRNAs with middle-expressed throughout lncRNA body region showed the highest methylation levels in CHH context in cassava. Both of low-expressed and non-expressed lncRNA displayed the same CHG and CHH methylation levels, which were lower than those of lncRNAs with high-expressed and middle-expressed (Fig. 2e, f). The results indicated that no correlation was observed between DNA CG methylation and lncRNA-expression levels, there was positive correlation for mCHG and mCHH methylation levels and lncRNA expression in 2x and 4x cassava.

Transposons with changed DNA methylation caused by WGD altered the expression of nearby PCGs and lncRNAs

The results of the transcriptome analysis indicated that 33,030 PCGs and 13,517 lncRNAs were detected in this study, which formed the dataset for the subsequent study. Comparison of expression level of PCGs and lncRNAs between 2x and 4x cassava indicated that only 359 PCGs and 402 lncRNAs were differentially expressed, respectively. That is, compared with 2x cassava, a large number of PCGs and lncRNAs were not significantly expressed in 4x cassava genotype that has obtained double alleles (Wilcoxon rank sum test, PCG P value = 0.5061817, n = 33,030; lncRNA P value < 0.003689, n = 13,517). Compared with 2x cassava, there were 173 PCGs up-regulated, and there were 186 PCGs down-regulated (Additional file 8: Fig. S6a); there were 204 lncRNAs up-regulated, and 198 lncRNAs down-regulated in 4x cassava (Additional file 8: Fig. S6b).

The differentially expressed PCGs were then used to be GO enrichment analysis. The 173 up-regulated PCGs in 4x cassava were involved in 'cellular process', 'metabolic process', 'response to stimulus', 'reproduction', etc. (Additional file 9: Fig. S7). The 186 down-regulated PCGs fell into 'metabolic process', 'cellular process', 'response to stimulus', 'development process', etc. (Additional file 9: Fig. S7).

To identify the two hypotheses mentioned above, the question is whether TEs may affect expression of neighboring PCGs and lncRNAs that were involved in WGD-induced variation in cassava. To address this question, 33,030 PCGs and 13,517 lncRNAs were used to examine whether they were inserted or surrounded by TEs. The results showed that approximately 48% of PCGs had TEs insertion into their bodies. Most of the TE insertions were DNA transposons, about 28% and 62% of PCGs were inserted by DNA transposons within bodies and 8-kb flanking regions, respectively (Fig. 3a). About 40% of lncRNAs had TEs into their bodies, 80% of lncRNAs were inserted by TEs within 2 kb, 94% lncRNAs overlapped with TEs within 4-kb flanking regions. Most of the TEs insertion into lncRNAs were retrotransposons, 25% and 52% of lncRNAs had retrotransposons insertion into their bodies and within 8-kb flanking regions, respectively (Fig. 3b). Moreover, we found that in 2x and 4x cassava, PCGs without neighboring TEs were expressed at higher levels than those inserted or surrounded by TEs (Fig. 3c). The expression levels of lncRNAs inserted or surrounded by TEs were higher than those without nearby TEs in the two cassava genotypes (Fig. 3d). The average PCG-expression level was positively correlated with the distance to the closest TE (2x P value = $4.36E-148$; 4x P value = $4.69E-156$) (Fig. 3e), for both 2x and 4x cassava, in contrast, the average lncRNA-expression level was negatively correlated with the distance to the closet TE (2x P value = $4.41E-25$; 4x P value = $1.80E-39$) (Fig. 3f). In addition, the average PCG-expression level was negatively correlated with the number of TEs within 4-kb flanking regions (2x P value = $6.38E-32$; 4x P value = $5.23E-33$) (Fig. 3g), however lncRNA-expression level was positively correlated with the number of TEs within 4-kb flanking regions in the two genotype cassavas (2x P value = 0.002433 ; 4x P value = 0.0007531) (Fig. 3h). Collectively, these results indicated that PCG-expression levels were suppressed by the abundance and physical distances from adjacent TEs, and lncRNA-expression levels were activated by the abundance and physical distances from proximal TEs.

Considering expression level of adjacent PCGs were negatively correlated with the state of methylated TEs in *Arabidopsis thaliana* and rice [21, 50], we compared DNA methylation of TEs from the whole genome, gene body, and flanking 4-kb regions between 2x and 4x cassava. Methylation of class II TEs, for both PCGs and lncRNAs, were lower than that of class I TEs for all three contexts (Fig. 4a, b), and methylation levels of TEs or class II TEs for PCGs in 4x cassava were higher than those of 2x except for flanking TEs in CHH context (Fig. 4a). Nevertheless, TEs or class I TEs inserted or surrounding lncRNAs from 2x cassava exhibited hypermethylation in CHG and CHH contexts (Fig. 4b).

We divided PCG-flanking regions into different bins and compared methylation levels between two TE classes located within them (Fig. 4c). Obviously, methylation levels of class II TEs nearby PCGs were lower than that of class I TEs in all three contexts. The CG methylation levels of class II TEs gradually decreased with increased distances from PCGs, and the valley of CG methylation levels of TEs appeared within 0.5-kb flanking regions. For mCHG and mCHH sites, class II TEs methylation levels gradually

increased with increased distances from PCGs. Notably, 4x cassava exhibited hypermethylation in CG and CHH contexts in PCG-flanking regions of class II TEs. To sum up, hypomethylation of class II TEs inserted in PCGs body regions in CHG and CHH contexts and hypermethylation of class II TEs neighboring PCGs in CG and mCHH contexts, may be a direct response factor to overcome genome-dosage effects following WGD in 4x cassava.

Parallely, we divided lncRNA-flanking regions into different bins and compared methylation levels between two TE classes located within them (Fig. 4d). Similarity with that of PCGs, methylation levels of class II TEs nearby lncRNA in all three contexts were lower than those of class I TEs. The profile of methylation levels of class I TEs did not show always rising for lncRNAs, which was different from that of PCGs. The CHG methylation levels of class I TEs had the same dynamic change with that of CHH methylation levels with increased distances from lncRNAs excluding within 0 to 0.5-kb flanking regions. The CG methylation levels of class I TEs appeared to be higher in 4x than that of 2x in 0~2.5-kb regions, however, CG methylation levels of class I TEs were lower in 4x in the flanking regions after 0~2.5-kb regions. Critically, in 4x cassava, class I TEs exhibited hypomethylation state in the CHG and CHH contexts compared with 2x in lncRNA-flanking regions, with the exception that CHG methylation state in ~1.5 to 3.0-kb lncRNA-flanking regions observed in 4x was the same as that of 2x, and this result was consistent with the genome-wide methylation level of the five types of class I TEs (Fig. 1d, e). Taking together, reduction of CHG and CHH methylation levels of class I TEs nearby lncRNAs in autotetraploid cassava may be a mechanism that suppressed expression of adjacent lncRNAs in autotetraploid cassava with double alleles from the diploid line responding to genome-wide lncRNA dosage effects after WGD.

Usually, retrotransposons of the *Gypsy* and *Copia* superfamilies are the major partitions of lncRNA in plant genome, especially for lincRNA [39, 40, 51]. Then, we calculated the percentage of the diversity classes of lncRNAs detected in this study, and found that lincRNAs (or called intergenic lncRNA) accounted for the largest proportion (42%) of the whole lncRNAs (Additional File 10: Fig. S8). We also found that the lncRNA loci contained more *Gypsy* and *Copia* segments than the other TEs at the exon and intron sequences together with 8-kb flanking regions in the cassava genome. *Gypsy* showed the largest proportion of lncRNA-overlapped TEs, followed by *Copia*, due to its largest share of TEs in the cassava genome (Additional File 11: Fig. S9). In order to understand whether the hypomethylated state of *Gypsy* or *Copia* was the effector overcoming genome-dosage effects in autotetraploid cassava, we depicted the correlation diagram of the DNA methylation level of *Gypsy* and *Copia* related to the distance from the closest lncRNA and lincRNA in 2x and 4x cassava, respectively (Fig. 4e-h). The results indicated that only the profile of methylation levels of *Gypsy* segment related to the closest lincRNA was found to be almost the same with that of class I TEs related to the closet lncRNA in all three contexts, with the exception that CHG methylation in 1.5~3.0-kb lncRNA-flanking regions observed in 4x was slightly higher than that of 2x, after all, CHG methylation levels of class I TEs in 4x cassava were the same as that of 2x in this region (Fig. 4e).

Comparisons of DMRs between 2x and 4x cassava

To investigate the potential effect of WGD, we identified a total number of 922 CG, 608 CHG and 51 CHH DMRs (Fig. 5a). 64.92% of mCG sites were hypermethylated while 43.09% of mCHG and 17.05% of mCHH sites exhibited hypomethylation in 4x cassava (Fig. 5a).

At the whole-genome level, most DMRs came from the mCG context, hardly any CHH hyper-DMRs existed in 4x cassava (Fig. 5b). The number of CG hyper-DMRs in 2x cassava was half that in 4x cassava, while the number of CHG-DMRs and CHH-DMRs in 2x is more than that in 4x (Fig. 5b). Genome-wide analysis showed that DMR of all three contexts were inclined to localize in intergenic and TEs regions rather than PCGs and lncRNAs, CHG-DMR within lncRNAs of 4x cassava increased in comparison with 2x cassava, there is no CHH hyper-DMR lncRNAs in 4x cassava (Fig. 5c). Analysis of PCG-flanking and lncRNA-flanking 4-kb regions revealed that the distribution of DMRs was similar to that throughout the whole genome (Fig. 5c).

Discussion

All flowering plants, such as rice, soybean (*Glycine max*) and cassava, etc., are paleo-polyploids that have undergone ancient WGD events during their evolution with the most recent one occurring 10 million years ago [52-56]. The common occurrence of WGD suggests an advantage for formed neo-polyploidy plants to enhance adaptation to stressful circumstances [27], consistent with the rapid expression divergence between duplicate genes [57-59]. Allopolyploid models have provided numerous clues to understand polyploidy formation, however, they can be confounded by the entanglement of both WGD and hybridization [60]. Conversely, autopolyploids have made changes via WGD, ruling out disturbances from incompatible genomes. Newly resynthesized polyploids, which can be induced by colchicine, have enabled biologists to provide insights into genomic changes that occur in response to autotetraploidization.

Large-scale variations of DNA methylation have been found in allopolyploid plants [2-3, 61-62]. Yin et al. [63] reported that CHG methylation variation is sensitive to regulate the stability of the genome of synthetic polyploidy rapeseeds (*Brassica napus* L.). The balanced genomic variations accompany pervasive convergent and concerted changes in DNA methylation and gene expression among *Arabidopsis* allopolyploids [64]. However, relative little is known about the effects of autopolyploidization. Since WGD stimulated genome doubling events with dosage balance, we asked whether this was attributed to purify simple noise arising from the chaos of genomic shock. In order to address this question, we examined PCG- and lncRNA-expression profiles, and genome-wide TEs methylation variation in autotetraploid cassava and its diploid line. Our findings showed that TEs methylation variations restrained expression of adjacent PCGs and lncRNAs, indicating that it is an effective way to overcome genome-dosage effects caused by WGD.

This work is the first study detected the lncRNA methylation feature in the cassava genome. There is no correlation observed between DNA CG methylation and lncRNA expression, which is obvious contrast between PCG body CG methylation and PCG-expression levels in cassava. There is negative correlation

between non-CG methylation and PCG-expression levels, however, positive correlation observed between non-CG methylation and lncRNA-expression levels. Our results indicated that WGD may have no influence over CG methylation of lncRNAs, and there is no correlation observed between CG methylation and lncRNA expression. These data were consistent with the result that the two cassava genotypes exhibited the same methylation levels of class I TEs in mCG sites in lncRNA-flanking regions.

TEs, nearly ubiquitous in lncRNAs, represent a major force shaping the lncRNA repertoire of plants and animals, through their capacity to move and spread in genomes in a lineage-specific fashion [65-66]. The TE sequences embedded in lncRNAs are critical for the biogenesis of lncRNAs and likely crucial for their function [67-73]. In our study, most of the insertion into lncRNAs were retrotransposons in the cassava genome, which was coincided with the results from previous studies [41, 51, 74-75]. Therefore, the impact of polyploidization on DNA methylation pattern of class I TEs partially reflected the impact of WGD on methylation pattern of lncRNAs. The methylation levels of lncRNAs are highly hypomethylated in the CHG and CHH contexts, partially reflecting hypomethylation patterns in the CHG and CHH contexts of genome-wide class I TEs and class I TEs inserted or surrounding lncRNAs in autotetraploid cassava. We also observed that the methylation level profile of *Gypsy* related to the closest lncRNA was highly similar to that of class I TEs related to the closest lncRNA. Therefore, methylation variation of *Gypsy* arising from class I TEs play an important role in suppressing expression levels of the nearby lncRNA in autotetraploid cassava after WGD. The impact of methylation variation of class I TEs on adjacent lncRNA-expression levels may be partially attributed to the effect of methylation variation of *Gypsy* on the expression of nearby lncRNAs in autotetraploid cassava responding to WGD.

The previous results from *Arabidopsis* and rice, the number, distance of TEs, and methylation levels of TE affected nearby PCG activity [21, 50, 76]. Our study further confirmed that although there were 28% PCGs inserted by class II TEs within their bodies, most of PCGs were inserted by class II TEs within 4-kb flanking regions (Fig. 1a). Class II TEs were preferential to localize in euchromatin regions near PCGs, and the number of class II TEs nearby PCGs is increased. Autotetraploid cassava displayed hypermethylation of class II TEs in CG context in PCG-flanking regions, which is consistent with increased genome-wide CG methylation level of class II TEs. Nevertheless, class I TEs are inclined to localize in heterochromatic regions, the distances of class I TEs were close to lncRNAs, resulting in an increased number of class I TEs near lncRNAs. Compared with diploid donor, autotetraploid cassava exhibited hypomethylation of class I TEs in CHG and CHH contexts in lncRNA-flanking regions, which is consistent with reducing genome-wide methylation levels in mCHG and mCHH sites in autotetraploid cassava. It is probably that autopolyploidization acts as an effector that may stimulate TE activities. After short-term generations, neoautopolyploids survivors may gradually conquer genome-dosage effects and adapt WGD through many mechanisms, one of which could be the TEs hypermethylation or hypomethylation that would have effects on nearby PCGs or lncRNAs. This route was a fitness trade-off between keeping TEs inactivated and suppressing proximal PCG- and lncRNA-expression levels. These results suggested that genome-wide epigenetic silencing through DNA methylation of TEs might play prominent roles in adapting to genomic shock following WGD.

In this study, PCG-expression levels were suppressed in autotetraploid cassava that have obtained double alleles, with an increase methylation levels of class II TEs nearby PCGs in CG and CHH contexts being observed. Instead, compared with the diploid rice, increasing methylation status of class II TEs in mCHG and mCHH sites restrained expression levels of nearby PCGs in an artificial synthesized rice with a relatively stable genome [21]. The methylation levels of class II TEs that surrounded PCGs in all three contexts were reduced in the colchicine-induced autopolyploid switchgrass with vegetative propagated for three years [24]. Therefore, we speculated that different crop species may adopt response factors with different methylation patterns in response to genome imbalance, which may be related with the stages after WGD. CHH methylation of TEs may be more sensitive than CG and CHG methylation in overcoming genome-dosage effect to contribute to genomic stability in autopolyploid genome in plant. Methylation level variation of TEs may be a beneficial strategy to help neoautopolyploids to conquer the early challenges caused by WGD in autopolyploid plants.

Conclusion

Our study conducted genome-wide single-base-resolution maps of methylomes and lncRNAomes to test whether significant DNA methylation alterations that occur in response to WGD can be found by comparing synthesized autotetraploid cassava with its donor parent. Our findings indicate that autotetraploid cassava exhibits widespread DNA methylation changes in TEs. In particular, genome-wide hypermethylation of DNA transposons in mCG and mCHH sites was found to restrain the expression levels of neighboring PCGs in autotetraploid cassava, resulting in similar expression levels for most of PCGs between autotetraploid and diploid cassava. The decreased methylation levels of retrotransposons in mCHG and mCHH sites, which partly attributed to reduction methylation of *Cypsy* neighboring lncRNAs in autotetraploid cassava, may be a mechanism that may suppress the expression levels of nearby lncRNA from its diploid parent.

This work highlighted that DNA methylation variation in TEs may be as a response factor to overcome genome-dosage effects caused by WGD.

Materials And Methods

Plant materials

An autotetraploid cassava line ($2n = 4x = 72$) was artificially created by the cultivar 'Xinxuan 048' ($2n = 2x = 36$) using aqueous colchicine solution. 0.1 g/L colchicine was applied to lateral buds of diploid stem for 48 h at 35 °C in the dark condition on the shaker, then chromosome counting and plant architecture screening were carried out for the first two generations. Stem-propagated plants from each diploid and autotetraploid cassava were sown and grown in plastic pots with a photoperiod of 16/8 h (day/night) in the greenhouse of Guangxi Academy of Agricultural Sciences (GXAAS). At ~60 d after planting, the fifth leave (counting from the top of the plant) of nine individual plants of each genotype were sampled. Three plants each biological replicate were included and each genotype has triplicates.

The collected leaves were immediately frozen in liquid nitrogen, and stored at -80°C until total DNA and RNA extractions were performed.

Methylation data analysis

DNA extraction

Genomic DNA was extracted according to a plant genomic DNA kit (Tiangen, China) following the manufacturer's instructions. After genomic DNAs were extracted from the samples, DNA concentration and integrity were detected by NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) and Agarose Gel Electrophoresis respectively.

Library construction, sequencing and data filtering

The DNA libraries for Bisulfite sequencing (BS-seq) were prepared. Briefly, genomic DNAs were fragmented into 100-300 bp by Sonication (Covaris, Massachusetts, USA) and purified with MiniElute PCR Purification Kit (QIAGEN, MD, USA). The fragmented DNAs were end repaired and a single "A" nucleotide was added to the 3' end of the blunt fragments. Then the genomic fragments were ligated to methylated sequencing adapters. Fragments with adapters were bisulfite converted using Methylation-Gold kit (ZYMO, CA, USA), unmethylated cytosine is converted to uracil during sodium bisulfite treatment. Finally, the converted DNA fragments were PCR amplified and sequenced using Illumina HiSeq™ 2500 by Gene Denovo Biotechnology Co. (Guangzhou, China).

To get high quality clean reads, the reads containing more than 10% of unknown nucleotides and low quality reads containing more than 40% of low quality (Q-value ≤ 20) bases were removed from the raw reads generated from Illumina HiSeq™ 2500.

TE annotation

TEs were annotated by running RepeatMasker (<http://repeatmasker.org>) against a cassava reference genome sequence (v6.1, <https://phytozome.jgi.doe.gov/pz/portal.html#>). In detail, Tandem repeats finder (TRF) [77] software was used to predict tandem repeats. Prediction method of Interpersed repeat was as following: i) Considering some repeat sequences often have specific sequence characteristics, such as LTR transposon, which is characterized by symmetric long terminal repeat at both ends, we predicted long terminal repeats (LTR) transposons through LTR_FINDER [78], Helitron transposon by Helitroscanner [79], MITE transposon by MITE-Hunter [80]. LINE by MGEScan-nonLTR [81]. ii) Since the repeat sequence has multiple copies in the genome, multiple copies of the repeat sequence in the genome can be found through mutual alignment within the genome sequence. First, three softwares PILER [82], RepeatScout [83] and RepeatModeler [84] were used to obtain preliminary *de novo* prediction results, and then sequences classified as DNA and LINE are extracted and merged into one file. The redundancy of the filtering sequence itself, and the filtering standard identity $> 90\%$. iii) Homology construction based on the principles of structure prediction (signature) and *de novo* (*de novo*), a repeat sequence database was constructed, which was combined with Repbase database as the final repeat sequence database. Then

RepeatMasker software was used to predict the repeat sequence of sequencing data based on the constructed repeat sequence database [85]. Collectively, information of the annotated TEs on each chromosome was listed in Additional File 12: Table S3, and a dataset of 12,592 TEs was used for further analysis (Additional File 13: Table S4).

Methylation level analysis

The obtained clean reads were mapped to cassava reference genome using BSMAP software [86] (version: 2.90) by default. We used a custom Perl script to call methylated cytosines (mC) and these methylated cytosines were tested with the correction algorithm described in Lister et al. [87]. The overall methylation levels were calculated using a BSMAP package script according to the ratio of reads (mC)/ [reads (mC) + reads (non-mC)]. The methylation level was calculated based on methylated cytosine percentage in the whole genome, in each chromosome and in different regions of the genome for each sequence context (CG, CHG and CHH). To assess different methylation patterns in different genomic regions, the methylation profile at flanking 2-kb regions and PCG (lncRNAs or TEs) body was plotted based on the average methylation levels for each window.

Analysis of differentially methylated regions

To determine the differentially methylated regions (DMRs) between 2x genotype and 4x genotype sample, the minimum read coverage to call a methylation status for a base was set to 4. DMRs for CG, CHG and CHH context according to different criteria: 1) For all C, numbers in a window ≥ 20 , absolute value of the difference in methylation ratio ≥ 0.2 , and $q \leq 0.05$; 2) For CG, numbers of GC in a window ≥ 5 , absolute value of the difference in methylation ratio ≥ 0.25 , and $q \leq 0.05$; 3) For CHG, numbers in each window ≥ 5 , absolute value of the difference in methylation ratio ≥ 0.25 , and $q \leq 0.05$; 4) For CHH, numbers in a window ≥ 15 , absolute value of the difference in methylation ratio ≥ 0.25 , and $q \leq 0.05$.

Enrichment analysis of differentially PCGs

We conducted Gene ontology (GO) enrichment analysis using the topGO v. 2.24 package in R (<https://bioconductor.org/packages/3.3/bioc/html/topGO.html>) for the differentially PCGs using a hypergeometric test with a corrected $P \leq 0.05$.

lncRNA-seq and data analysis

Transcriptome sequencing was performed with the same leaf tissues used for BS-seq for each of the 2x and 4x genotype. Total RNA was extracted from 100 mg of tissue using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. The whole-transcriptome library construction and sequencing were performed by Gene Denovo Biotechnology Co. (Guangzhou, China). All the libraries were sequenced on an Illumina HiseqTM4000. After removing sequences containing adapters, poly-N and low quality reads, clean reads were aligned to the cassava reference genome (v6.1, <https://phytozome.jgi.doe.gov/pz/portal.html#>) by HISAT2 (version 2.1.0) with “-rna-strandness RF” and

other parameters set as a default [88]. The reconstruction of transcripts was carried out using Stringtie (version 1.3.4), which together with HISAT2 [89-90]. To identify novel genes, all of the reconstructed transcripts were mapped to the cassava reference genome and were divided into 12 categories by using Cuffcompare [91]. Gene abundances were quantified using RSEM (v 1.2.19) [92] and PCG- and lncRNA-expression levels were normalized using FPKM (Fragments Per Kilobase of transcript per Million reads). Two softwares CPC (version 0.9-r2) and CNCI (version 2) were used to assess the protein-coding potential of novel transcripts by default parameters [93-94]. The intersection of both non protein-coding potential results and non protein annotation results were chosen as lncRNAs.

Differential expression analysis of PCGs and lncRNAs was performed by DESeq2 [95] software between two different groups and by edgeR [96] between two samples. We used a false discovery rate (FDR) < 0.05 and fold change ≥ 2 as the thresholds to determine significant differences in PCG and lncRNA expression.

Abbreviations

WGD: Whole-genome duplication

PCG: Protein coding gene

TEs: Transposable elements

lncRNA: Long non coding RNA

LincRNA: Intergenic lncRNA

DMR: Differential methylation region

GXAAS: Guangxi academy of agricultural science

BS-seq: Bisulfite sequencing

TRF: Tandem repeats finder

LTR: Long terminal repeats

mC: Methylated cytosines

GO: Gene ontology

FPKM: Fragments per kilobase of transcript per million reads

FDR: False discovery rate

Declarations

Conflict of interest

The authors confirm no conflict of interest.

Author contributions

HY and LX planned and designed the research. LX, WZ, LL and XS analyzed the data. SC participated in the preparation of the materials, cultivation and management of the plants. LX wrote the paper and HY revised the paper. All authors commented on the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The data that supports the findings of this study are available in the supplementary materials of this article. All the data have been deposited in BioProject under accession numbers PRJNA728761.

Funding

This work was supported by National Key Research and Development Program of China (grant no. 2019YFD1000500), National Natural Science Foundation of China (grant no. 31960440), Guangxi Natural Science Foundation (grant no. 2020GXNSFAA259085) and Key Projects of Guangxi Natural Science Foundation (grant no. 2020GXNSFDA297031).

Acknowledgements

We thank Dr. X. Zhou (Chinese Academy of Tropical Agricultural Sciences), and Professor L. Chen and Dr. J. Song (Guangxi University) for critically suggestions. We thank the Guangzhou Gene denovo biotechnology Co., Ltd for assisting in the sequencing and data analysis.

References

1. Comai L. The advantages and disadvantages of being polyploidy. *Nature Reviews Genetics*. 2005; 6(11):836–846.
2. Madlung A, Wendel JF. Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet Genome Research*. 2013; 140(2-4):270–285.
3. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. *Genome Biology*. 2016; 17:37.
4. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis D, Soltis P, et al. Evolutionary genetics of genome merger and doubling in plants. *Annual Review Genetic*. 2008; 42:443–461.
5. Salmon A, Ainouche ML. Polyploidy and DNA methylation: new tools available. *Molecular Ecology*. 2010; 19(2):213–215.

6. Zhang K, Wang XW, Cheng F. Plant polyploidy: origin, evolution, and its influence on crop domestication. *Horticultural Plant Journal*. 2019; 5(6): 231–239.
7. Ramsey J, Schemske DW. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review Ecology and Systematics*. 1998; 29:467–501.
8. Xu Y, Zhong L, Wu X, Fang X, Wang J. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta*. 2009; 229: 471–483.
9. Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics*. 2010; 186(3): 801–812.
10. Havananda T, Brummer EC, Doyle JJ. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany*. 2011; 98(10):1633–1646.
11. Liu BB, Sun GL. microRNAs contribute to enhanced salt adaptation of the autopolyploid *Hordeum bulbosum* compared with its diploid Ancestor. *The Plant Journal*. 2017; 91(1):57–69.
12. Ramsey J, Schemske DW. Neopolyploidy in flowering plants. *Annual Review Ecology System*. 2002; 33:589–639.
13. Soltis DE, Soltis PS, Tate JA. Advances in the study of polyploidy since Plant speciation. *New Phytologist*. 2003; 161(1):173–191.
14. Soltis PS, Soltis DE. The role of hybridization in plant speciation. *Annual Review Plant Biology*. 2009; 60:561–588.
15. Parisod C, Holderegger R, Brochmann C. Evolutionary consequences of autopolyploidy. *New Phytologist*. 2010; 186(1):5–17.
16. Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*. 2016a; 210(2):391–398.
17. Martelotto LG, Ortiz JPA, Stein J, Espinoza F, Quarin CL, Pessino SC. A comprehensive analysis of gene expression alterations in a newly synthesized *Paspalum notatum* autotetraploid. *Plant Science*. 2005; 169(1):211–220.
18. Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyng S, et al. Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics*. 2007; 176(4):2055–2067.
19. Yu Z, Haberer G, Matthes M, Rattei T, Mayer K, Gierl A, et al. Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(41):17809–17814.
20. Allario T, Brumos J, Colmenero-Flores JM, Tadeo F, Froelicher Y, Talon M, et al. Large changes in anatomy and physiology between diploid Rangpur lime (*Citrus limonia*) and its autotetraploid are not associated with large changes in leaf gene expression. *Journal of Experiment Botany*. 2011; 62(8):2507–2519.
21. Zhang J, Liu Y, Xia EH, Yao QY, Liu XD, Gao LZ. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proceedings of*

- the National Academy of Sciences of the United States of America. 2015a; 112:7022–7029.
22. Yin L, Qu JJ, Zhou HW, Shang XH, Fang H, Lu J, et al. Comparison of leaf transcriptomes of cassava “Xinxuan 048” diploid and autotetraploidy plants. *Genes & Genomics*. 2018; 40:927–935.
 23. Xiao L, Shang XH, Cao S, Xie XY, Zeng WD, Lu LY, et al. Comparative physiology and transcriptome analysis allows for identification of lncRNAs imparting tolerance to drought stress in autotetraploid cassava. *BMC Genomics*. 2019; 20:514.
 24. Yan HD, Bombarely A, Xu B, Wu BC, Frazier TP, Zhang XQ, et al. Autopolyploidization in switchgrass alters phenotype and flowering time via epigenetic and transcription regulation. *Journal of Experiment Botany*. 2019; 70(20):5673–5686.
 25. Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. *Nature Reviews Genetics*. 2013; 14:471–482.
 26. Song Q, Chen ZJ. Epigenetic and developmental regulation in plant polyploids. *Current Opinion Plant Biology*. 2015; 24:101–109.
 27. Chen ZJ. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review Plant Biology*. 2007; 58:377–406.
 28. Li ZH, Lu X, Gao Y, Liu SJ, Tao M, Xiao H, et al. Polyploidization and epigenetics. *Chinese Science Bulletin*. 2011; 56:245–252.
 29. Rival A, Ilbert P, Labeyrie A, Torres E, Doulebeu S, Personne A, et al. Variations in genomic DNA methylation during the long-term in vitro proliferation of oil palm embryogenic suspension cultures. *Plant Cell Report*. 2013; 32:359–368.
 30. Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature*. 2009; 461:423–426.
 31. Saze H, Kakutani T. Differentiation of epigenetic modifications between transposons and genes. *Current Opinion Plant Biology*. 2011; 14(1):81–87.
 32. Ibarra CA, Feng XQ, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*. 2012; 337(6100):1360–1364.
 33. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*. 2007; 39(1):61–69.
 34. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Review Genetic*. 2010; 11(3):204–220.
 35. Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *The Plant Cell*. 2001; 13(8):1749–1759.
 36. Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids.

- Plant Physiology. 2002; 129(2):733–746.
37. Feschotte C, Jiang N, Wessler SR. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics*. 2002; 3:329–341.
 38. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genetics*. 2013; 9(4):e1003470.
 39. Wang D, Qu ZP, Yang L, Zhang QZ, Liu ZH, Do T, et al. Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *The Plant Journal*. 2017; 90(1):133–146.
 40. Wang X, Ai G, Zhang CL, Cui L, Wang JF, et al. Expression and diversification analysis reveals transposable elements play important roles in the origin of *Lycopersicon*-specific lncRNAs in tomato. *New Phytologist*. 2016; 209(4):1442–1455.
 41. Zhao T, Tao XY, Feng SL, Wang LY, Hong H, Ma W, et al. lncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biology*. 2018; 19:195.
 42. Raven P, Fauquet C, Swaminathan MS, Borlaug N, Samper C. Where next for genome sequencing? *Science*. 2006; 311(5760):468–468.
 43. FAO. 2010. FAOSTAT. <http://www.fao.org/faostat>. Accessed on August 18, 2010.
 44. Carvalho R, Guerra M. Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. *Hereditas*. 2002; 136(2):159–168.
 45. Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, et al. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Theoretical and Applied Genetics*. 2003; 107:1083–1093.
 46. Siqueira MVB, Pinheiro TT, Borges A, Valle TL, Zatarim M, Veasey EA. Microsatellite polymorphisms in cassava landraces from the Cerrado biome, Mato Grosso do Sul, Brazil. *Biochemistry Genetic*. 2010; 48:879–895.
 47. Zhou HW, Zeng WD, Yan HB. In vitro induction of tetraploids in cassava variety 'Xinxuan 048' using colchicine. *Plant Cell, Tissue and Organ Culture*. 2017; 128:723–729.
 48. Wang HF, Beyene G, Zhai JX, Feng SH, Fahlgren N, Taylor NJ, et al. CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112(44):13729–13734.
 49. Yan HD, Bombarely A, Xu B, Frazier TP, Wang CR, Chen PL, et al. siRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass. *Biotechnology for Biofuels*. 2018; 11:208.
 50. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*. 2009; 19(8):1419–1428.
 51. Lv YD, Hu FQ, Zhou YF, Wu FL, Zhou L, Gaut BS. Transposable elements contribute to regulatory hubs stress-related long noncoding RNAs in maize. *The Plant Journal*. 2019a. DOI: 10.1101/801704.

52. Jiao YN, Wickett N, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011; 473:97–100.
53. Jiang WK, Liu YL, Xia EH, Gao LZ. Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. *Plant Physiology*. 2013; 161(4):1844–1861.
54. Barker MS, Husband BC, Pires JC. Spreading wings and flying high: the evolutionary importance of polyploidy after a century of study. *American Journal of Botany*. 2016b; 103(7):1139–1145.
55. Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nature Review Genetic*. 2017; 18(7):411–424.
56. Ren R, Wang HF, Guo C, Zhang N, Zeng LP, Chen YM, et al. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Molecular Plant*. 2018; 11(3):414–428.
57. Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology*. 2006; 7:R13.
58. Ha M, Kim ED, Chen ZJ. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(7):2295–2300.
59. Ezoe A, Shirai K, Hanada K. Degree of functional divergence in duplicates is associated with distinct roles in plant evolution. *Molecular Biology and Evolution*. 2020; 38(4):1447–1459.
60. Qin JX, Mo RR, Li HX, Ni ZF, Sun QX, Liu ZS. The transcriptional and splicing changes caused by hybridization can be globally recovered by genome doubling during allopolyploidization. *Molecular Biology and evolution*. 2021; 38(6):2513.
61. Salmon A, Ainouche ML, Wendel JF. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (*Poaceae*). *Molecular Ecology*. 2005; 14(4):1163–1175.
62. Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiology*. 2006; 140(1):336–348.
63. Yin LQ, Zhu ZD, Huang LJ, Luo X, Li Y, Xiao CW, et al. DNA repair- and nucleotide metabolism-related genes exhibit differential CHG methylation patterns in natural and synthetic polyploids (*Brassica napus* L.). *Horticultural Research*. 2021; 8:142.
64. Jiang XY, Song QX, Ye WX, Chen ZJ. Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. *Nature Ecology and Evolution*. 2021; DOI:[10.1038/s41559-021-01523-y](https://doi.org/10.1038/s41559-021-01523-y).
65. Davis MP, Carrieri C, Saini HK, van Dongen S, Leonardi T, Bussotti G, et al. Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. *EMBO Report*. 2017; 18:1231–1247.

66. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology*. 2012; 13(11):R107.
67. Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, et al. Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. *New Phytologist*. 2016; 210(4):1298–1310.
68. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009; 326(5956):1112–1115. □
69. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*. 2015b; 33(5):531–537.
70. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 2012; 492(7429):423–427.
71. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*. 2014; 46:567–572.
72. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genetics*. 2010; 42(12):1113–1117.
73. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, et al. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*. 2011; 147(4):773–788.
74. Golicz AA, Singh MB, Bhalla PL. The long intergenic noncoding RNA (lincRNA) landscape of the soybean genome. *Plant Physiology*. 2018; 176(3):2133–2147.
75. Lv YD, Hu FQ, Zhou YF, Wu FL, Gaut BS. Maize transposable elements contribute to long non-coding RNAs that are regulatory hubs for abiotic stress response. *BMC Genomics*. 2019b; 20(1):864.
76. Wang X, Weigel D, Smith LM. Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genetics*. 2013; 9:e1003255.
77. Benson G. Tandem Repeats Finder: a program to analyze DNA sequences. *Nucleic Acids Research*. 1999; 27:573–580.
78. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*. 2007; 35:W265–W268.
79. Xiong W, He L, Lai J, Dooner HK, Du C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(28):10263–10268.
80. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Research*. 2010; 38(22):e199.

81. Lee H, Lee M, Ismail WM, et al. MGEScan: a Galaxy-based system for identifying retrotransposons in genomes. *Bioinformatics*. 2016; 32(16):2502–2504.
82. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics*. 2005; 21 Suppl 1(Suppl 1):i152–8.
83. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*. 2005; 21(Issue suppl_1):i351–i358.
84. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*. 2002; 12(8):1269–1276.
85. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*. 2009; 25, chapter 4: unit 4.10.
86. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics*. 2009; 10:1.
87. Lister R, Pelizzola M, Dowen RH, David Hawkins R, Hon G, Tonti-Filippini JL, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462:315.
88. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 2013; 14(4):R36.
89. Pertea M, Pertea GM, Antonescu CM, Chang T, Mendell J, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*. 2015; 33(3):290–295.
90. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*. 2016; 11:1650–1657.
91. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*. 2012; 7(3):562–578.
92. Li B, Dewey CN. RSEM accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12:323.
93. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei LP, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*. 2007; 35(Issue suppl_2):W345–349.
94. Sun L, Luo HT, Bu DC, Zhao GG, Yu KT, Zhang CH, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*. 2013; 41(17):e166.
95. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014; 15(12):550.
96. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–140.

Additional File

Figures

Figure 1

DNA methylation patterns in PCGs, lncRNAs and TEs. Average methylation level distributions over PCGs (a), lncRNAs (b), and TEs (c). Average methylation level distribution over class I TEs (d, e) and class II TEs (f, g). The average level for each 100-bp interval is plotted. The dashed lines for the PCG, lncRNA and TE regions indicate the transcriptional start (left) and end (right) sites.

Figure 2

Association between DNA methylation level and gene expression in 2x and 4x cassava. Effect of DNA methylation of mCG (a), mCHG (b) and mCHH (c) on genome-wide PCGs expression. Effect of DNA methylation of (d) mCG, (e) mCHG and (f) mCHH on global lncRNAs expression.

Figure 3

TEs altered expression of neighboring PCGs and lncRNAs in cassava. The percentage of PCGs (a) and lncRNAs (b) inserted by TEs in their bodies and within 2-kb flank regions. The expression levels of PCGs (c) and lncRNAs (d) inserted by TEs or not. "+" means TEs inserted in this region; *P value < 0.05; **P value < 0.01. PCGs (e) and lncRNAs (f) expression level related to the distance to the closest TE. "0" indicates genes overlapped with TEs in body regions. Error bars indicates SEM. The expression levels of PCGs (g) and lncRNAs (h) related to the number of neighboring TEs.

Figure 4

DNA methylation level related to the distance from the closest PCG and lncRNA. DNA methylation of TEs from the PCG (a) and lncRNA (b) body of whole genome, together with flanking 4-kb regions. TE methylation level related to the distance from the closest PCG (c) and lncRNA (d). Gypsy (e), and Copia (f) methylation level related to the distance from the closest lncRNA. Gypsy (g), and Copia (h) methylation level related to the distance from the closest lncRNA.

Figure 5

DMR distribution. (a) DMR numbers of all three contexts and proportions of hyper- and hypo-methylated regions in all three contexts for 4x cassava compared with 2x. (b) Genome coverage and number of defined hypermethylation DMRs. (c) Distribution of hypermethylation DMRs in the whole-genome (WG) and gene-flanking 4-kb regions (GF).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.jpg](#)
- [Fig.S2.pdf](#)
- [Fig.S3.pdf](#)
- [Fig.S4.pdf](#)
- [Fig.S5.pdf](#)
- [Fig.S6.pdf](#)
- [Fig.S7.jpg](#)
- [Fig.S8.pdf](#)
- [Fig.S9.pdf](#)
- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.xls](#)