Brief communication

Title: Haplotype resolved DNA methylome of African cassava genome

Zhenhui Zhong[1], Suhua Feng[1,2], Ben N. Mansfeld[3], Yunqing Ke[1], Weihong Qi[4], Yi-Wen Lim[5], Wilhelm Gruissem[5], Rebecca S. Bart[3], Steven E. Jacobsen[1,2,6†]

[1] Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA

[2] Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California, Los Angeles, CA 90095, USA

[3] Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

[4] Functional Genomics Center Zurich, ETH Zurich and University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

[5] Institute of Molecular Plant Biology, Department of Biology, ETH Zürich, Universitätsstrasse 2, 8092 Zürich, Switzerland

[6] Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095, USA

† Correspondence to: jacobsen@ucla.edu (S.E.J.)

Keywords: Haplotype resolved, DNA methylation, Cassava

Cytosine DNA methylation is involved in biological processes such as transposable element (TE) silencing, imprinting, and X chromosome inactivation. Plant methylation is mediated by MET1 (mammalian DNMT1), DRM2 (mammalian DNMT3), and two plant-specific DNA methyltransferases, CMT2 and CMT3 (Law and Jacobsen, 2010). *De novo* DNA methylation in plants is established by DRM2 via the plant specific RNA-directed DNA methylation (RdDM) pathway that depends on two DNA-dependent RNA polymerases, Pol IV and Pol V (Gallego-Bartolome *et al.*, 2019; Law and Jacobsen, 2010; Stroud *et al.*, 2013). The DNA methylome of cassava has been previously documented based on its haploid collapsed genome (Wang *et al.*, 2015). Since the cassava genome is highly heterozygous, DNA methylome analysis of the haplotype-collapsed genome misses many features of the methylome. With the development of long read sequencing and chromosomal conformation capture techniques, haplotype resolved genomes are available for highly heterozygous genomes (Mansfeld *et al.*, 2021; Qi *et al.*, 2022; Zhou *et al.*, 2020), which provides high-quality reference genomes facilitating studies of haplotype resolved DNA methylomes.

To dissect the haplotype resolved DNA methylome of cassava, we conducted whole genome methylome studies in two haplotype genome resolved accessions of cassava

(TME7 and TME204) using whole-genome bisulfite sequencing (WGBS) and enzymatic methyl-seq (EM-seq), respectively (Feng *et al.*, 2020; Mansfeld *et al.*, 2021; Qi *et al.*, 2022). Sequencing reads were mapped to different haplotypes individually allowing zero mismatches and one best hit, which allowed the separation of reads belonging to different haplotypes. Overall, we found that the two haplotypes have similar whole genome methylation levels (Figure 1A, Supplemental Figure 1). We further plotted methylation levels over transcribed and flanking 1000 bp regions of protein-coding genes and transposable elements, and observed similar methylation levels between different haplotypes (Figure 1B,C, Supplemental Figure 2AB).

Previous studies have revealed large numbers of haplotype-specific structural variants (SVs) in cassava (Mansfeld *et al.*, 2021; Qi *et al.*, 2022). To understand how DNA methylation is associated with these SVs, we analyzed methylation levels of these SVs. Focusing on places where haplotype-specific SVs occur, we found that flanking regions of SVs have approximately two times lower methylation levels than random controls and SVs, suggesting that SVs preferentially take place at lowly methylated regions, with the SVs turning these regions to heavily methylated regions (Figure 1D-F). Allele-specific expression (ASE) refers to preferential expression of the allele transcribed from one haplotype (Gaur *et al.*, 2013). ASEs can be cataloged into complete ASEs and partial ASEs. Since the number of partial ASEs and non-ASE controls are much higher than that of complete ASEs, we selected an equal number of complete ASEs and partial ASEs or non-ASE controls with 100 times iterations. Interestingly, complete ASE genes showed higher gene body CHG methylation than partial (p = 0.05) and non-ASE genes and partial ASE genes showed lower CG methylation than others (p = 3.8e-08, Figure 1G).

Next, we analyzed differentially methylated cytosines between haplotypes (haplotypic DMCs). We first aligned the two haplotypes and identified syntenic cytosines (Supplemental Figure 2C). Methylation levels of syntenic cytosines were compared between two haplotypes. We observed at least three scenarios (Figure 1H) that resulted in differential methylations between haplotypes: 1) SNP/InDel in one haplotype leads to the loss of the cytosine (47.17%) among which more than 99.25% are SNP variants (Figure 1I; see TME204 data in Supplemental Figure 2D); 2) Cytosine context alterations (e.g. CG in hap1 and CHH in hap2) that leads to the methylation level changes detected in our analysis (28.98%); and 3) Cytosines stay in the same contexts between haplotypes but exhibit different methylation levels (23.85%). For DMCs caused by scenarios 2 and 3, the majority of them were CHH variations (Figure 1I). When CG sites are mutated into CHG or CHH sites, they are more likely to be hypomethylated; whereas CHH sites are more likely to be hypermethylated when mutated into CHG and CG sites (Figure 1JK, Supplemental Figure 2EF). Furthermore, we found that DMCs are frequently detected at TEs and distal intergenic regions, while depleted at gene body and flanking regions, suggesting that TEs are hot spots for frequent DNA methylation changes (Figure 1L, Supplemental Figure 2G). Finally, we investigated the genetic diversity of DMC sites and 400 bp flanking sequences and

found that nucleotide diversity of DMC sites are significantly higher than that of flanking sequences (Figure 1MN), demonstrating that DMC sites are under more frequent natural selection. Together, our analyses compared haplotype resolved DNA methylomes of cassava, and place genomic heterozygosity within the haplotypic epigenetic regulatory landscape.

Conflicts of interest

The authors declare no conflict of interest.

Authors' contributions

Z.Z., S.F., B.M., W.G., R.B. and S.J. conceived and designed the research. Z.Z., W.Q., and Y.K. performed data analysis. Y.L. and S.F. prepared sequencing libraries. Z.Z., S.F., and S.J. wrote the manuscript.

Data availability statement

The DNA methylome data are available in Gene Expression Omnibus under accession number GSE192748 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192748).

References

Feng, S., Zhong, Z., Wang, M., and Jacobsen, S.E. (2020) *Efficient and accurate determination of genome-wide DNA methylation patterns in Arabidopsis thaliana with enzymatic methyl sequencing. Epigenetics and Chromatin*, **13**.

Gallego-Bartolome, J., Liu, W., Kuo, P.H., Feng, S., Ghoshal, B., Gardiner, J., et al. (2019) *Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in Arabidopsis. Cell*, **176**, 1068-1082. e19.

Gaur, U., Li, K., Mei, S., and Liu, G. (2013) *Research progress in allele-specific expression and its regulatory mechanisms. J Appl Genet*, **54**, 271–283.

Law, J.A. and Jacobsen, S.E. (2010) *Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics*, **11**, 204–220.

Mansfeld, B.N., Boyher, A., Berry, J.C., Wilson, M., Ou, S., Polydore, S., et al. (2021) *Large structural variations in the haplotype‐resolved African cassava genome. The Plant Journal*, **108**, 1830–1848.

Qi, W., Lim, Y.-W., Patrignani, A., Schläpfer, P., Bratus-Neuenschwander, A., Grüter, S., et al. (2022) *The haplotype-resolved chromosome pairs of a heterozygous diploid African cassava cultivar reveal novel pan-genome and allele-specific transcriptome features. Gigascience*, **11**.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y. v, and Jacobsen, S.E. (2013) *Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell*, **152**, 352–364.

Wang, H., Beyene, G., Zhai, J., Feng, S., Fahlgren, N., Taylor, N.J., et al. (2015) *CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proceedings of the National Academy of Sciences*, **112**, 13729–13734.

Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J.P., et al. (2020) *Haplotype-resolved genome analyses of a heterozygous diploid potato. Nat Genet*, **52**, 1018–1023.

Figure 1. Haplotype resolved DNA methylome of African cassava.

(A) Whole genome methylation levels of hap1 and hap2 haplotypes in TME7 (n=3) and TME204 (n=1). (B) Metaplot of CG, CHG, and CHH methylation levels over protein-coding genes and flanking 1Kb sequences of hap1 and hap2 haplotypes in TME204. (C) Metaplot of methylation levels over transposon elements and flanking 1Kb sequences in TME204. (D-F) Metaplot of methylation levels over structural variation regions and flanking 1Kb sequences in TME7. Equal length regions were randomly selected in the genome (blue line). (G) Methylation levels of complete, partial, and non-ASE genes in TME7. (H) Representative screenshots of DMCs. Methylation site losses and mC context changes are indicated on top of the track (red: hap1; blue: hap2). (I) Ratio of DMCs caused by SNP/InDel, where the cytosines are lost, and DMCs caused by different methylation levels. (J) Numbers of different types of cytosine context variations between the two haplotypes. (K) Consensus sequences of DMCs. (L) Genomic distribution enrichment of DMCs. (M-N) Nucleotide diversity of 400bp flanking regions of DMC in TME7 (M) and TME204 (N). Panels I, J, L, and K show results for TME7. Results for TME204 are shown at Supplemental Figure 2.