




Brief Communication

Multi-omics mapping of chromatin interaction resolves the fine hierarchy of 3D genome in allotetraploid cotton

Xianhui Huang^{1,†}, Xuehan Tian^{1,†}, Liuling Pei¹, Xuanxuan Luo¹, Yuqi Zhang¹, Xingtang Zhang², Xianlong Zhang¹ , Longfu Zhu^{1,*}  and Maojun Wang^{1,*} 

¹National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, China

²Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

Received 11 April 2022;

revised 30 May 2022;

accepted 16 June 2022.

*Correspondence (Tel +86-027-87283955; fax +86-027-87280196; email lfzhu@mail.hzau.edu.cn (L.Z.) and (Tel +86-027-87283955; fax +86-027-87280196; email mjwang@mail.hzau.edu.cn (M.W.))

[†]These authors contribute equally to this study.

Keywords: Pore-C, Hi-C, ChIA-PET, 3D genome, cotton.

Allotetraploid cotton (*Gossypium hirsutum*) is cultivated globally as an important fibre and oil crop, and also serves as a model for genomic analysis of allopolyploidization. The cotton research is currently dedicated to identifying functional genome components, including protein-coding genes and non-coding regulatory elements, to guide the improvement of agronomic traits (Huang *et al.*, 2021). The research is also extended to understand the 3-dimensional (3D) genome structure, building on the increasing evidence of its role in gene transcription and regulation (Domb *et al.*, 2022). However, the understanding of higher-order chromatin structure and long-range transcriptional *cis*-regulatory elements remains fragmentary. In this study, three chromatin architecture mapping technologies, including Pore-C, Hi-C and ChIA-PET, were used to resolve the fine hierarchy of 3D genome in *G. hirsutum*. In combination with ATAC-Seq, ChIP-Seq and RNA-Seq, this study addresses the putative effect of chromatin interaction on homoeologous gene transcription between two subgenomes.

Limited by experimental principles, Hi-C does not perform well for exploring multiple interactions (Lieberman-Aiden *et al.*, 2009). However, Pore-C can complement this shortcoming by intracellular cross-linking, *in situ* digestion (DpnII), *in situ* proximity ligation, reverse cross-linking and Nanopore long-read sequencing (Figure 1a; Deshpande *et al.*, 2022). To construct a fine 3D genome map of cotton, we produced an ultra-high-resolution (3 Kb) Hi-C library (233 Gb) and a Pore-C library (113 Gb) with appropriate resolution (20 Kb) using leaves of *G. hirsutum* (Figure 1b, Tables S1 and S2). The length of the A compartment (100 Kb) accounts for 44.34% and 43.02% of the entire genome identified by Pore-C and Hi-C, respectively, with a high consistency of eigenvectors and high overlap (92.99%) between them (Figure 1c and d). There are 7702 and 8028 topologically associating domains (TADs) identified from Hi-C and Pore-C at the 20 Kb resolution, respectively, with high consistency of insulation score (Figure 1e and f). These results indicate that Pore-C library can be used to identify A/B compartment and TAD architecture accurately.

Compared with Hi-C, the prominent advantage of Pore-C is to detect multiple interactions. Our result shows that Pore-C identifies 43.56% of multiple interactions (order ≥ 3) (Figure 1g). To investigate the multiple interactions between large-scale domains in cotton, we identified TAD cliques that represent clusters of interacting TADs. Compared with the observation in Hi-C data, Pore-C revealed more complex interactions (Figure 1h and i). Meanwhile, more genes were participated in higher levels of TAD clique interactions in Pore-C (Figure 1j). By analysing the proportion of A/B compartment in TAD cliques and its distribution on the chromosome, we found that the larger TAD cliques included more B compartments, closer to the centromere (Figure 1k and l). We further found that the genes in larger TAD cliques were more likely to be down-regulated, which suggests they may be located in a more inhibitory micro-environment (Figure 1m). These results indicate that Pore-C represents an efficient approach for mapping higher-order chromatin structure in terms of completely investigating multi-interactions over large domains, which helps to understand the status and intranuclear position of functional genomic components.

We next used the Hi-C, ChIA-PET (H3K4me3) and ChIP-Seq (H3K4me3, H3K27ac and H3K27me3) data to study the chromatin loops and their potential role in gene transcription (Table S3). We identified 31 047 gene–gene (G–G) loops that linked two genes together, 40 035 gene–non-coding regions (G–N) loops and 121 415 other loops (N–N) at the 3 Kb resolution. It is found that genes located in G–G loop anchors had higher expression levels, associated with higher content of active histone modifications (Figure 1n, Figure S1). Subsequently, to investigate the regulatory role of non-coding regions on gene expression bridged by loops, we identified 4670 and 3076 putative long-range transcriptional *cis*-regulatory elements (CREs) by combining ATAC-Seq, ChIP-Seq and chromatin loops. These CREs were associated with H3K4me3, H3K27ac and H3K27me3 modifications, and were overlapped with Tn5 transposase-hypersensitive sites (THSs), linked with 2317, 2713, 4193 and 6683 genes, respectively. The genes modified by H3K27me3 had the lowest expression levels, possibly associated with the inactive role of H3K27me3 (Figure 1o). These data provide a resource for characterizing functional regulatory elements in non-coding genomic regions.

Using the fine 3D genome mapping of allotetraploid cotton, we further investigated the putative effect of different 3D genome architectures of two subgenomes on transcriptional regulation. We found 9115 homoeologous gene pairs with different spatial locations on TADs, that is one located in TAD boundary and the other in TAD interior. Compared with randomly selected gene pairs, this gene set contained a significantly higher

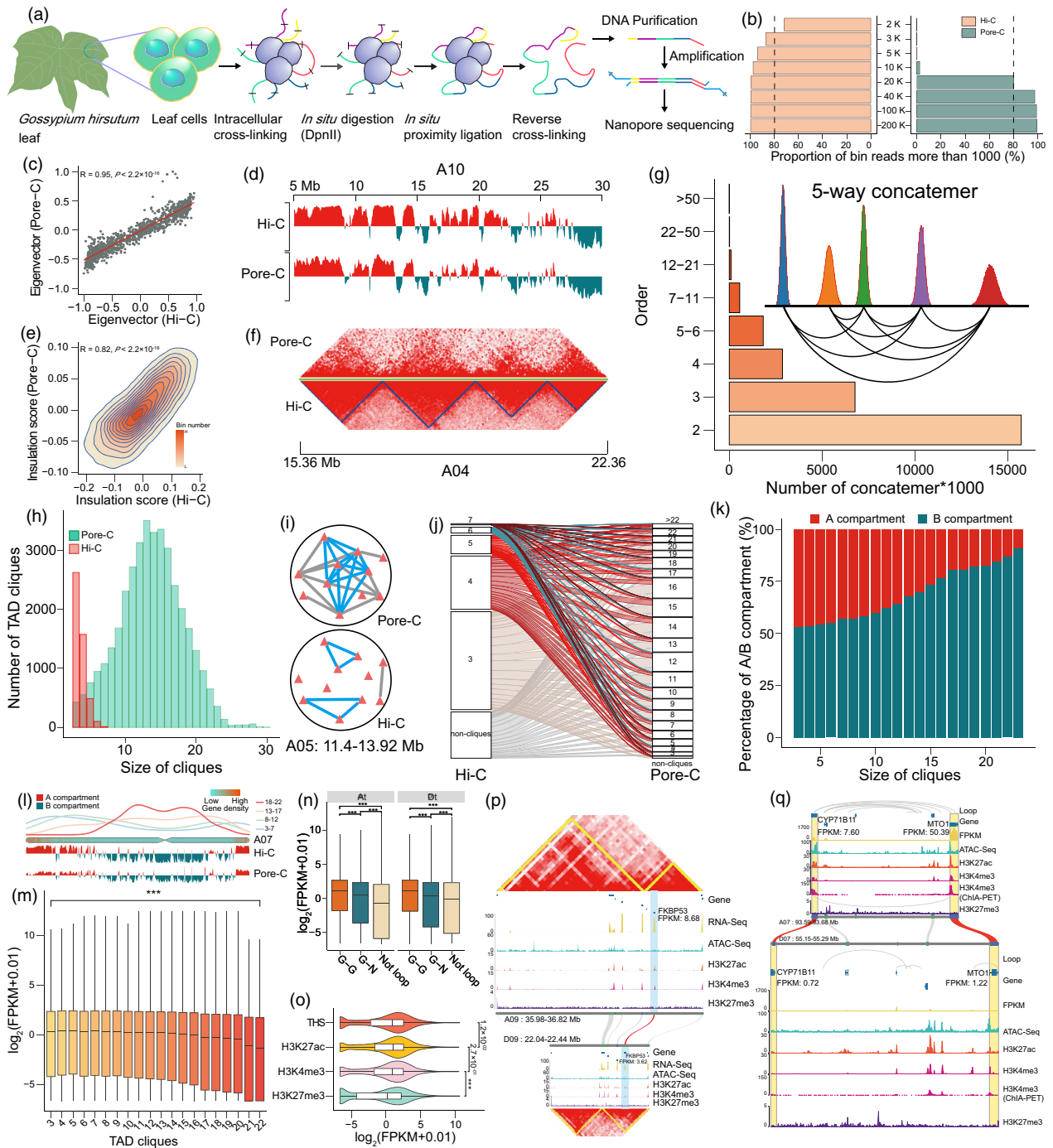


Figure 1 Multi-omics mapping of the fine 3D genome structure and implications in transcriptional regulation in allotetraploid cotton. (a) Schematic of the Pore-C method. (b) The bar plot shows the resolutions of Hi-C and Pore-C. (c) The point plot represents the consistency of eigenvectors between Hi-C and Pore-C. (d) One representative genome region shows the status of A/B compartment identified by Hi-C and Pore-C. (e) The density plot shows the consistency of insulation score between Hi-C and Pore-C. (f) The heatmap shows TAD structures at the 20 Kb resolution. (g) The histogram shows the number of concatemers and the schematic representation of 5-way concatemer interactions identified by Pore-C. (h) The bar plot shows the number of TAD cliques identified from Pore-C and Hi-C libraries. (i) The network plot represents the interactions of TADs. (j) The snake plot shows the distribution of genes in different TAD cliques between Pore-C and Hi-C. (k) The bar plot shows the ratio of A and B compartments in different TAD cliques. (l) The distribution of TAD cliques and A/B compartment on chromosomes. (m) The expression level of genes located in different TAD cliques. (n) The expression level of genes linked by different loops. (o) The expression level of genes regulated by different CREs. (p) Biased expression of homoeologous genes with topological change from TAD interior in the Dt subgenome to the boundary in the At subgenome. (q) Biased expression of homoeologous genes with differential loops between subgenomes.

proportion exhibiting subgenomic expression bias (1088 versus 424, Pearson's chi-squared test <0.001), which may suggest different spatial locations of homoeologous genes were associated with expression imbalance (Figure 1p). Meanwhile, an analysis of CREs showed that 5977 homoeologous gene pairs were regulated by subgenome-differential CREs through chromatin loops, and 716 gene pairs exhibited expression bias towards one of two subgenomes, which was also a significant higher proportion than random gene pairs (716 versus 41, Pearson's chi-squared test <0.001 ; Figure 1q). These results indicate that homoeologous gene expression was not only influenced by the difference in the distribution of TAD positions but also influenced by the changed CREs linked by regulatory loops. Analysis of the relationship between the change of subgenomic 3D architecture and homoeologous gene expression bias contributes to the understanding of subgenomic regulatory divergence after allopolyploidization in cotton.

To sum up, we used multi-omics to comprehensively probe the fine 3D genomic architecture of allotetraploid cotton. With the Pore-C technique, we investigated the characteristics of complex high-level multiple interactions over large domains and the potential effects on gene expression. We also revealed that biased expression of homoeologous gene pairs was implicated with the subgenome-asymmetric organization of higher-order chromatin architecture. This study highlights the necessity of decoding the hierarchy of 3D genome for a wider understanding of complicated transcriptional regulation in polyploid plants.

Acknowledgements

This study is financially supported by the National Key Research and Development Program of China (2021YFF1000900) and the National Natural Science Foundation of China (31922069, 32170645). This study is also supported by the Foundation of Hubei Hongshan Laboratory (2021hszd014). We thank the high-performance computing platform at National Key Laboratory of Crop Genetic Improvement in Huazhong Agricultural University.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

M.W. and L.Z. designed the experiments and managed the project. X.H. and X.T. analysed the data. X.T., L.P., X.L. and Y.Z.

performed the experiments. Xingtian Z. and Xianlong Z. contributed to project discussion. X.H. wrote the manuscript draft, M.W., Xianlong Z. and L.Z. revised it.

Materials and methods

Refer to [Material S1](#).

Data availability statement

All data are publicly available in the NCBI (<https://www.ncbi.nlm.nih.gov/>) under BioProject numbers PRJNA824233 and PRJNA576032.

References

- Deshpande, A.S., Ulahannan, N., Pendleton, M., Dai, X., Ly, L., Behr, J.M., Schwenk S. *et al.* (2022) Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-022-01289-z>
- Domb, K., Wang, N., Hummel, G. and Liu, C. (2022) Spatial features and functional implications of plant 3D genome organization. *Annu. Rev. Plant Biol.* **73**, 173–220.
- Huang, G., Huang, J.Q., Chen, X.Y. and Zhu, Y.X. (2021) Recent advances and future perspectives in cotton research. *Annu. Rev. Plant Biol.* **72**, 437–462.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit I. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 The box plot shows the proportion of active histone modifications (H3K27ac and H3K4me3) in different types of loop anchors.

Figure S2 The Venn plot shows the overlap number of peaks identified from ChIA-PET and ChIP-Seq libraries.

Table S1 The sequencing statistics of Hi-C library.

Table S2 The sequencing statistics of Pore-C library.

Table S3 The sequencing statistics of ChIA-PET library.

Material S1 Plant materials.