Telomere-to-telomere pear (*Pyrus pyrifolia*) reference genome reveals segmental and whole genome duplication driving genome evolution

https://academic.oup.com/hr/article/10/11/uhad201/7308738

Published on < Horticulture Research> IF: 8.7

ADVANCING GENOMICS
IMPROVING LIFE

Case Study

A telomere-to-telomere (T2T) genome is crucial for gaining profound insights into genome evolution and optimizing crop enhancement. A paper titled "Telomere-to-telomere pear (*Pyrus pyrifolia*) reference genome reveals segmental and whole genome duplication driving genome evolution", presents the first complete, gap-free pear genome assembly by integrating Illumina, Pacific Biosciences (PacBio), Oxford Nanopore Technology (ONT), and high-throughput chromosome conformation capture (Hi-C) sequencing data. Leveraging this high-quality genome assembly, the researchers conducted comprehensive analyses to characterize genome duplication events and the divergence of duplicated genomic regions. The insights gained from these analyses enhance the understanding of the pear genome's structural composition and the functional diversification of duplicated genes.

Research Background

Advancements in long-read sequencing technologies have facilitated the assembly of T2T genomes using PacBio HiFi reads, ONT ultra-long reads, and Hi-C data. T2T plant genomes, including *Arabidopsis*, rice, and maize, have been successfully assembled. These genomes provide detailed representations of complex sequences, allowing for the exploration of genetic variations and repetitive sequences in previously unclear regions.

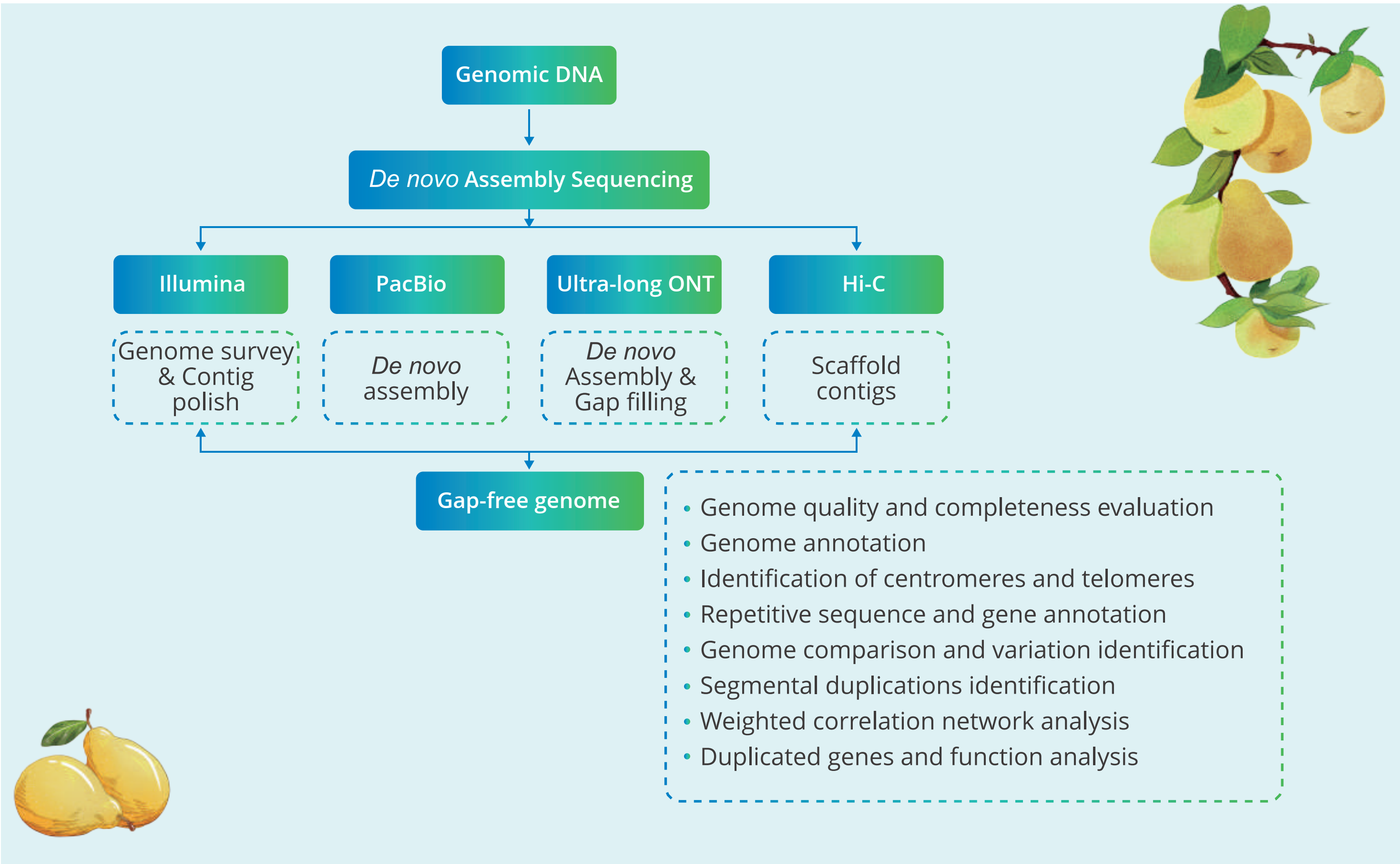
While several pear genome assemblies have been released, these incomplete versions frequently lack fine-scale genetic details, particularly within duplicated regions, thereby limiting our comprehension of gene duplication and subsequent sub-/neo-functionalization events.

Research Pipeline

To assemble the genome, first, Illumina reads were used for genome survey to estimate the genome size and heterozygosity levels. Next, both PacBio HiFi reads and ultra-long ONT reads were employed for *de novo* assembly, with the Illumina data used to polish the assembled contigs derived from ONT reads. The genome assembled using PacBio HiFi data was selected as the reference genome. The ONT-assembled genome was then merged with the HiFi-assembled reference to fill in the remaining gaps. Finally, Hi-C data was employed to correct and scaffold contigs.

To evaluate the quality of the assembly, clean sequencing reads were mapped to each haplotype, then genome coverage and mapping rate were calculated. Genome completeness was evaluated using BUSCO and CEGMA.

The structure of protein-coding genes and gene functions were predicted. Genome comparisons between this assembly and two other assemblies of *Pyrus pyrifolia* were performed.



Genomic DNA

De novo Assembly Sequencing

Illumina

PacBio

Ultra-long ONT

Hi-C

Genome survey & Contig polish

De novo assembly

De novo Assembly & Gap filling

Scaffold contigs

Gap-free genome

- Genome quality and completeness evaluation
- Genome annotation
- Identification of centromeres and telomeres
- Repetitive sequence and gene annotation
- Genome comparison and variation identification
- Segmental duplications identification
- Weighted correlation network analysis
- Duplicated genes and function analysis

Results

- By incorporating Illumina, PacBio HiFi, ONT ultra-long, and Hi-C sequencing, the first T2T gap-free genome for pear was generated. The assembled genome size is 501.20 Mb with all 34 telomeres and 17 gap-free T2T pseudomolecules identified and reconstructed. The 17 centromeric regions were predicted to arrange in sizes from 1.35 to 2.80 Mb. Furthermore, 99.00% core genes were evaluated as complete based on the BUSCO assessment.

- This study identified 53.94 Mb of segmentally duplicated (SD) regions (10.76% of the assembled pear genome). The 1,531 duplicated gene pairs within SDs were enriched for stress response pathways. Notably, disease resistance genes were exclusively found in intrachromosomal SDs. These findings suggest SDs enhance the pear's capacity for environmental adaptation and stress tolerance.

- This study revealed that 7,495 gene pairs exhibited asymmetrical expression patterns, suggesting divergence in their functional roles. Moreover, 876 duplicated gene pairs in the pear genome appear to have undergone processes of sub-/neo-functionalization. Duplicated genes that are associated with important agronomic traits can serve as resources for pear genetic improvement.

- The transcription factors MYB10 and MYB114 are essential regulators of anthocyanin biosynthesis in pear. The study confirmed that these two TFs are present in multiple copies within the pear genome, implying that the duplicated gene copies have been preserved through a mechanism of reduced expression levels to achieve dosage balance.

Conclusion

This paper presents the first T2T gap-free pear genome assembly, offering a precise and comprehensive view of the pear genome's structural composition and gene functionality landscape.

Reference

Sun, M., Yao, C., Shu, Q., He, Y., Chen, G., Yang, G., Xu, S., Liu, Y., Xue, Z., & Wu, J. (2023). Telomere-to-telomere pear (*Pyrus pyrifolia*) reference genome reveals segmental and whole genome duplication driving genome evolution. Horticulture research, 10(11), uhad201. https://doi.org/10.1093/hr/uhad201

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A. J., Abramson, B. W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., Yelina, N., Hartwick, N., Colt, K., Smith, L. M., Ton, J., Kakutani, T., Martienssen, R. A., Schneeberger, K., Lysak, M. A., Berger, F., ... Henderson, I. R. (2021). The genetic and epigenetic landscape of the Arabidopsis centromeres. Science (*New York, N.Y.*), 374(6569), eabi7489. https://doi.org/10.1126/science.abi7489

Song, J. M., Xie, W. Z., Wang, S., Guo, Y. X., Koo, D. H., Kudrna, D., Gong, C., Huang, Y., Feng, J. W., Zhang, W., Zhou, Y., Zuccolo, A., Long, E., Lee, S., Talag, J., Zhou, R., Zhu, X. T., Yuan, D., Udall, J., Xie, W., ... Chen, L. L. (2021). Two gap-free reference genomes and a global view of the centromere architecture in rice. *Molecular plant*, 14(10), 1757–1767. https://doi.org/10.1016/j.molp.2021.06.018

Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., Li, T., Hu, J., Wang, K., Wang, C., Xin, B., Zhao, H., Song, W., Hufford, M. B., Schnable, J. C., Jin, W., & Lai, J. (2023). A complete telomere-to-telomere assembly of the maize genome. *Nature genetics*, 55(7), 1221–1231. https://doi.org/10.1038/s41588-023-01419-6

Follow us on LinkedIn

Novogene Corporation Inc.

2921 Stockton Blvd. Suite 1810, Sacramento CA 95817

☎ 916-252-0068 ext. 383 ✉ inquiry_us@novogene.com www.novogene.com