Plant Communications
Correspondence

CellPress
Partner Journal

A telomere-to-telomere genome assembly for greater yam (*Dioscorea alata*)

Dear Editor,

Dioscorea alata L. (2n = 2x = 40, 3x = 60, and 4x = 80) is a perennial, tangled, herbaceous vine in the Dioscoreaceae family that migrated from Asia to Africa and America via the Indian Ocean and the Caribbean (Sharif et al., 2020). The rich starch and crude protein in its inflated tuber (commonly known as the greater yam) make this plant an essential crop in the tropical and subtropical regions of Southeast Asia, Africa, and Latin America (Sharif et al., 2020). Moreover, *D. alata*, *D. cayenensis* and *D. rotundata* represent the three major food crops among 11 cultivated yam species, whereas the other eight are often referred to as "minor yams" (Lebot et al., 2023). The global production of yam tubers reached 75 million tons in 2021 (Food and Agriculture Organization Statistical Database, 2022), indicating their importance in the food supply. A high-quality reference genome serves as an essential resource for the molecular breeding of crops to improve their quality and resistance to pathogens. With the recent publication of a chromosome-level genome, efforts have been made to identify the genes/loci associated with quality traits and resistance to pathogens in greater yam (Bredeson et al., 2022; Mota et al., 2024). For example, a population genetic and quantitative trait locus (QTL) mapping study based on a reference genome revealed three anthracnose resistance QTLs, including two containing typical nucleotide-binding site leucine-rich repeat (NLR)-type resistance proteins (Bredeson et al., 2022). However, the published reference genome assembly still contains over 500 gaps (Bredeson et al., 2022), raising the possibility of previously unidentified NLR genes and other important agronomic genes in the gap regions. Addressing these gaps is therefore key to obtaining a more comprehensive gene annotation of the greater yam. In this study, we present a telomere-to-telomere (T2T) gap-free genome of the greater yam cultivar 'Ziyushenshu', providing a complete reference genome for future research.

To achieve a T2T gap-free genome assembly of the greater yam cultivar 'Ziyushenshu', we first used 25 Gb of PacBio high-fidelity (HiFi) long reads to generate an assembly of 508 Mb, comprised of 38 contigs (Supplemental Methods). High-resolution chromosome conformation capture (Hi-C) data was then used to assist in scaffolding, and all contigs were successfully anchored to 20 scaffolds, corresponding to 20 pseudochromosomes (Supplemental Figure 1). A total of 155 Gb of Nanopore reads from the Oxford Nanopore Technologies platform, including 18.7 Gb of ultra-long reads, were used to fill the gaps in the assembly. Instead of merely extracting ultra-long reads that could span genomic gaps, we assembled the Nanopore reads *de novo* using a different methodology and obtained several Nanopore assemblies (Supplemental Methods). By comparing the anchored HiFi+Hi-C assembly with the Nanopore assemblies, we discovered that all 18 gaps of the HiFi+Hi-C assembly could be compensated by the Nanopore assemblies, resulting in a gap-free genome of greater yam with a total size of 511.5 Mb (Figure 1A). Our assembly is larger than that previously published (480 Mb) (Bredeson et al., 2022) and more consistent with estimates derived from flow cytometry and *k*-mer analyses (Supplemental Figure 2; Supplemental Table 1). Both the PacBio HiFi and Nanopore reads were mapped to the assembly at extremely high rates (99.99% for both), providing evidence for the high fidelity and completeness of our assembled greater yam genome. The telomeres and centromeres of all 20 pseudochromosomes were identified (Figure 1B). The telomeres range from 469 bp to 16.7 kb in length, with an average length of 4.7 kb, featuring the characteristic TTTAGGG sequence repeating 677 times (Supplemental Table 2), whereas the centromeres range from 105.3 to 6854.3 kb in length, with an average length of 1786.717 kb (Supplemental Table 3).

BUSCO (embryophyta_odb10) estimates for the greater yam genome assembly and annotation captured 97.78% and 95.3% of intact genes, respectively. Annotation identified 62.58% of the genome as repetitive sequences. A total of 34 816 DNA transposons and 103 304 retroelements represented 4.59% and 23.27% of the entire genome, respectively. By incorporating extensive RNA sequencing data (Supplemental Methods) from multiple organs/tissues, 30 781 protein-coding genes were annotated in the newly assembled genome, with 5 574 more than in the previously published *D. alata* genome of an African cultivar (Bredeson et al., 2022).

Given that the distribution of synonymous substitution sites (Ks) for anchor paralogs shows two similar signature peaks (~1.00 and 1.5) for all three *Dioscorea* species long before speciation (divergence peaks < 0.5), the three *Dioscorea* species is expected to have experienced two rounds of shared whole-genome duplications (Figure 1C; Supplemental Figure 3) (Sun et al., 2022). Hence, it is more likely that the seemingly "doubled" chromosome number between *D. zingiberensis* (2n = 20) and *D. rotundata*/*D. alata* (2n = 40) results from chromosome fissions/fusions rather than independent whole-genome duplications (Tamiru et al., 2017). Using the maximal parsimony method and collinearity comparison, the common ancestor of the three species was reconstructed to possess 19 haploid chromosomes, and the karyotype evolution of *Dioscorea* was deduced accordingly (Figure 1D; Supplemental Figures 4 and 5; Supplemental Note 1; Supplemental Data 1). The *D. rotundata–D. alata* lineage maintained all ancestral karyotypes except for one single fission event (Figure 1D). This event split one ancestral chromosome into two (chromosomes 5 and 16), thus resulting in 20 chromosomes for each haplotype. In contrast, *D. zingiberensis* in the other lineage underwent significant fusion events, including three non-homologous chromosome fusions, four end-to-end chromosome

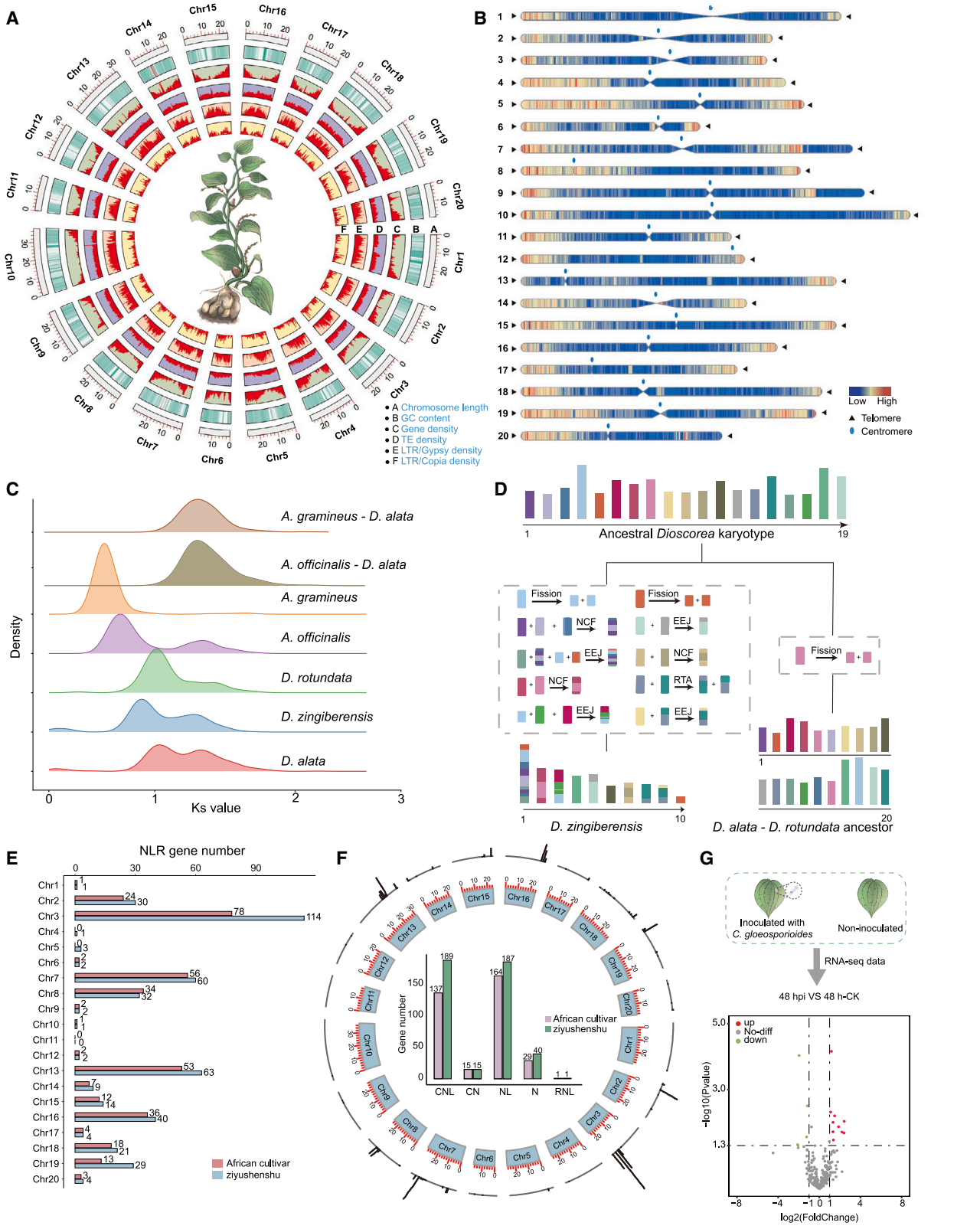

Figure 1. T2T gap-free assembly of *D. alata*.

(A) Circos plot showing the T2T genome assemblies for the *D. alata* cultivar 'Ziyushenshu'. The tracks from A to F display the following: chromosome length, guanine+cytosine (GC) content, transposable element (TE) density, gene density, long terminal repeat (LTR)/Gypsy density, and LTR/Copia density.

(legend continued on next page)

joins, and one reciprocal translocation, ultimately resulting in a reduction to 10 chromosomes.

The cultivar 'Ziyushenshu' sequenced in this study is notable for its resistance to *C. gloeosporioides*. NLR proteins act as major intracellular immune receptors of plants in response to various pathogens (Ngou et al., 2022). Therefore, identifying the full NLR repertoire serves as a fundamental resource for exploring functional resistance genes against various pathogens. However, the repetitive nature of NLR genes—with encoding domains capable of tolerating duplications and deletions of entire repeats—and their tendency to form gene clusters on chromosomes present significant challenges for the comprehensive and correct assembly and annotation of these crucial receptors of innate immunity in plants (Meyers et al., 2003). Using the acquired T2T genome, we systematically identified NLR family genes in this *D. alata*-resistant cultivar (Supplemental methods), recognizing a total of 432 NLR genes (Figure 1E; Supplemental Figure 6). The NLR genes identified in the *D. alata* 'Ziyushenshu' T2T genome exceeded those in the previously published *D. alata* African cultivar genome by 100, with 4 NLR genes located in centromeres and gap regions (Figure 1E; Supplemental Table 4). Notably, an overwhelming majority (92. 8%) of 'Ziyushenshu' NLR genes are organized into gene clusters on chromosomes, leaving only 32 NLR genes as singletons (Figure 1F; Supplemental Table 4). Additionally, a greater number of intact NLR genes (190), encoding complete CC (coiled-coil)/RPW8 (resistance to powdery mildew 8), NBS (nucleotide-binding site), and LRR (leucine-rich repeat) domains, were annotated in the T2T genome (Figure 1F; Supplemental Table 4), compared to the 138 intact genes in the African cultivar. The significantly enhanced annotation comprehensiveness and accuracy of this super gene family also reflect the necessity of a high-quality T2T genome and its potential importance in molecular breeding for resistant cultivars (Liu et al., 2023). By analyzing transcriptomic data from leaves of the *D. alata* cultivar 'Ziyushenshu' before and after inoculation with *C. gloeosporioides*, we identified 10 NLR genes significantly upregulated in response to pathogen invasion (Figure 1G; Supplemental Tables 5 and 6). These genes are located on chromosomes 3 (four genes), 7, 8, 12, and 13 (three genes). A BLASTp search analysis against over 400 known functional NLR resistance genes reveals that the eight *C. gloeosporioides*-induced 'Ziyushenshu' NLRs show the highest amino acid sequence similarity to Hrt1, RPS5, SUMM2, Sr46, PvLOV, and RPi-abpt, respectively (Supplemental Table 7). Notably, although several NLR genes (Supplemental Figure 7) were located in the QTLs on chromosomes 9 and 15 of the African cultivar, conferring potential anthracnose resistance (Bredeson et al., 2022), none of their orthologs in the 'Ziyushenshu' cultivar were significantly upregulated during *C. gloeosporioides* infection (Supplemental Figure 8). These results suggest that the two *D. alata* cultivars may have adopted different loci to combat anthracnose. Alternatively, the differences in the induced members of the NLR gene family may be due to the different strains of *C. gloeosporioides* used in the two studies. A phylogenetic analysis was conducted on NLR genes from the 'Ziyushenshu' T2T genome and African cultivar genome, along with those from three other species within Dioscoreaceae. The results show that four of the *C. gloeosporioides*-induced NLRs were recently duplicated in *D. alata*, including two NLRs specifically duplicated in 'Ziyushenshu' (Supplemental Figure 9). These results further suggest the differentiation of NLR repertoires among closely related species and even different cultivars within the same species. Due to the lack of a mature system for functional characterization, the candidate genes cannot be verified experimentally. Nevertheless, the breeding of greater yams may consider these potential NLR genes, allowing future production results to test our inferences. Recently, a T2T genome of a *D. alata* cultivar from the West Indies was posted as a preprint (Dossa et al., 2025). This cultivar has pure white tubers, distinct from our 'Ziyushenshu' with purple tubers. The future molecular breeding of *D. alata* may benefit from the integration of diverse genomic resources.

DATA AVAILABILITY

The PacBio HiFi, ultra-long Nanopore, Hi-C, and next-generation sequencing data, transcriptomes, genome assembly, and annotation files have been deposited at the China National GeneBank Database under accession number CNP0006270. Details of the China National GeneBank Database Run IDs are provided in Supplemental Table 8.

FUNDING

This work was supported by the National Natural Science Foundation of China (32172089 and 32270241), the Jiangsu Key Laboratory for the Research and Utilization of Plant Resources (JSPKLB202207), the National Administration of Traditional Chinese Medicine High-level Key Discipline Construction Project (grant no. zyyzdxk-2023293) and Fundamental Research Funds for the Central Universities from Nanjing Agricultural University. Z.-Q.S. is supported by the Outstanding Young Teacher of the "QingLan Project" of Jiangsu Province.

ACKNOWLEDGMENTS

No conflicts of interest declared.

(B) Identification of telomeres and centromeres in the genome of the *D. alata* cultivar 'Ziyushenshu'. A heatmap on each chromosome shows gene density, with a gradient scale from red to blue indicating high to low gene density, respectively.
(C) Polyploidy events in *D. rotundata*, *D. alata*, *D. zingiberensis*, *A. gramineus*, and *A. officinalis.*
(D) Evolutionary trajectory of karyotypes in three *Dioscorea* species. Different colors represent different ancestral chromosomes. Chromosome structural variations are highlighted in blue, where fission indicates chromosome breakages and fusion indicates potential chromosome connections. Nested chromosome fusion (NCF), end-to-end joining (EEJ), and reciprocal translocation of chromosome arms (RTA) are also indicated.
(E) Comparative profile of NLR genes between the genomes of the *D. alata* cultivar 'Ziyushenshu' and the African cultivar.
(F) Distribution of NLR singletons and clusters anchored to chromosomes in the genome of the *D. alata* cultivar 'Ziyushenshu', with a comparative analysis of domain composition against the NLRs in the African cultivar.
(G) Transcriptomic response of NLR genes in the genome of the *D. alata* cultivar 'Ziyushenshu' following *C. gloeosporioides* inoculation. Red and green dots represent significantly upregulated and downregulated NLR genes, respectively.

AUTHOR CONTRIBUTIONS

J.-Y.X., Z.-Q.S., X.-Q.S., and Y.-M.Z. designed the research. Z.-Y.W. and C.-A.Y. performed the genome assembly and annotation. Z.-Y.W., C.-A.Y., X.-Y.F., Y.-M.Z., Y.W., and S.-X.L. conducted the evolutionary analysis. Y.-M.Z. provided the reagents and materials used in this study. Z.-Y.W. and Y.-M.Z. wrote the manuscript. J.-Y.X., Z.-Q.S., and X.-Q.S. revised the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: September 25, 2024
Revised: October 31, 2024
Accepted: March 31, 2025
Published: April 2, 2025

Yan-Mei Zhang[1,2,5], Zhi-Yan Wei[3,5], Cheng-Ao Yang[3], Xing-Yu Feng[4], Yue Wang[1,2], Sai-Xi Li[4], Xiao-Qin Sun[1,2,], Zhu-Qing Shao[4,*] and Jia-Yu Xue[3,*]*

[1]Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China
[2]Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Nanjing 210014, China
[3]College of Horticulture, Bioinformatics Center, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China
[4]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China
[5]These authors contributed equally to this article.
***Correspondence: Xiao-Qin Sun (xiaoqinsun@cnbg.net), Zhu-Qing Shao (zhuqingshao@nju.edu.cn), Jia-Yu Xue (xuejy@njau.edu.cn)**
https://doi.org/10.1016/j.xplc.2025.101326

REFERENCES

Bredeson, J.V., Lyons, J.B., Oniyinde, I.O., Okereke, N.R., Kolade, O., Nnabue, I., Nwadili, C.O., Hribová, E., Parker, M., Nwogha, J., et al. (2022). Chromosome evolution and the genetic basis of agronomically important traits in greater yam. Nat. Commun. **13**:2001.

Dossa, K., Su, X., Zoclanclounon, Y.A.B., Chen, Y., Nawaz, M.A., Ni, P., Qiu, T., Chen, G., Malédon, E., Gravillon, M.C., et al. (2025). Haplotype-resolved telomere-to-telomere genome assembly of a male *Dioscorea alata* cultivar provides insights into sex determination mechanismREPRINT (Version 1). Research Square. https://doi.org/10.21203/rs.3.rs-5627324/v1.

Lebot, V., Lawac, F., and Legendre, L. (2023). The greater yam (*Dioscorea alata* L.): A review of its phytochemical content and potential for processed products and biofortification. J. Food Compos. Anal. **115**:104987.

Liu, Y., Zhang, Y.M., Tang, Y., Chen, J.Q., and Shao, Z.Q. (2023). The evolution of plant NLR immune receptors and downstream signal components. Curr. Opin. Plant Biol. **73**:102363.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell **15**:809–834.

Mota, A.P.Z., Dossa, K., Lechaudel, M., Cornet, D., Mournet, P., Santoni, S., Lopez, D., and Chaïr, H. (2024). Whole-genome sequencing and comparative genomics reveal candidate genes associated with quality traits in *Dioscorea alata*. BMC Genom. **25**:248.

Ngou, B.P.M., Ding, P., and Jones, J.D.G. (2022). Thirty years of resistance: Zig-zag through the plant immune system. Plant Cell **34**:1447–1478.

Sharif, B.M., Burgarella, C., Cormier, F., Mournet, P., Causse, S., Van, K.N., Kaoh, J., Rajaonah, M.T., Lakshan, S.R., Waki, J., et al. (2020). Genome-wide genotyping elucidates the geographical diversification and dispersal of the polyploid and clonally propagated yam (*Dioscorea alata*). Ann. Bot. **126**:1029–1038.

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X., and Liu, J. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol. Plant **15**:1841–1851.

Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., Yoshida, K., Uemura, A., Oikawa, K., Abe, A., et al. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. BMC Biol. **15**:86.