

BRIEF COMMUNICATION OPEN ACCESS

A Near Telomere-To-Telomere Genome Assembly and Graph-Based Pangenome of Tartary Buckwheat (*Fagopyrum tataricum*)

Wendy Li¹ | Hanfei Liang² | Jilin Sun² | Xiao Zhang¹ | Qiang He² | Peng Zhou² | Dongao Huo^{1,3} 

¹College of Biological Sciences and Technology, Taiyuan Normal University, Taiyuan, China | ²State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China | ³Modern Agricultural Development Center, Jincheng Agricultural and Rural Bureau, Jincheng, China

Correspondence: Peng Zhou (pzhou@caas.cn) | Dongao Huo (hzauidha@163.com)

Received: 31 March 2025 | **Revised:** 10 July 2025 | **Accepted:** 10 August 2025

Funding: This work was supported by Biological Breeding–National Science and Technology Major Project, 2023ZD04076. Critical Talent Workstation Project, TYSGJ202201. Basic Research Program of Shanxi Province, 202203021222244.

Keywords: comparative population genomics | gap-free genome | graph-based pan-genome | Tartary buckwheat

Tartary buckwheat (*Fagopyrum tataricum*), a member of the Polygonaceae family, is a crop of substantial medicinal and nutritional importance. Since the initial genome assembly (~450 Mb) was published in 2017, two subsequent versions have been reported (Zhang et al. 2017; Du and Liang 2019; He et al. 2023). However, these assemblies remain incomplete with persistent gaps. Moreover, current genomic studies have primarily focused on single nucleotide polymorphisms (SNPs), while structural variations (SVs)—a major contributor to genetic diversity—remain underexplored (Garg et al. 2024). Here, we present a near complete T2T genome assembly for the high-yield cultivar Heifeng1 of Tartary buckwheat. We complemented this reference genome with deep resequencing of 163 *F. tataricum* accessions and generated chromosome-level assemblies for 10 representative accessions (Figure 1A and Figure S1; Table S1). By integrating the T2T genome with these 10 new assemblies and two previously published chromosomal assemblies, we constructed the first graph-based pangenome for Tartary buckwheat.

The T2T assembly of *F. tataricum* cv. Heifeng1 incorporated multiple sequencing technologies: 51.8× PacBio HiFi reads (43.5 Gb), 95.8× Oxford Nanopore Technologies (ONT) ultra-long reads (23.5 Gb), 58.7× Illumina short reads (26.6 Gb), and 71.1× Hi-C

data (32.3 Gb) (Figure S2 and Table S2). Our assembly strategy yielded a gap-free near T2T assembly spanning 453.9 Mb across eight chromosomes, containing eight centromeres and 15 telomeres (Figure 1B; Figures S3 and S4; Table S3). Remapping of all sequencing data after quality control showed > 99.7% alignment rates with uniform coverage distribution across all chromosomes (Figures S5 and S6). Benchmarking metrics of the Heifeng1-T2T genome demonstrated exceptional assembly quality, with a contig N50 of 56.89 Mb, LTR Assembly Index (LAI) score of 14.76, 99.1% complete embryophyte BUSCOs, and a Merquy-estimated quality value (QV) of 50.59 (Table S4). Comparative analysis with existing assemblies revealed significant improvements in our genome, particularly in telomere resolution (Figure 1B). Notably, we identified and validated through Hi-C data a 5.4 Mb pericentric inversion on chromosome 7 (19.5–24.9 Mb; Figure S7), demonstrating the assembly's power to resolve complex structural variation.

Repetitive elements constituted 59.57% of the Heifeng1-T2T genome, with long terminal repeat (LTR) retrotransposons representing the most abundant class at 39.23% (Table S4). Tandem Repeat Sequences accounted for 3.784%, while protein-coding transposable elements (TEs) comprised 12.13% (Table S4).

Wendy Li and Hanfei Liang contributed equally to this article.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.

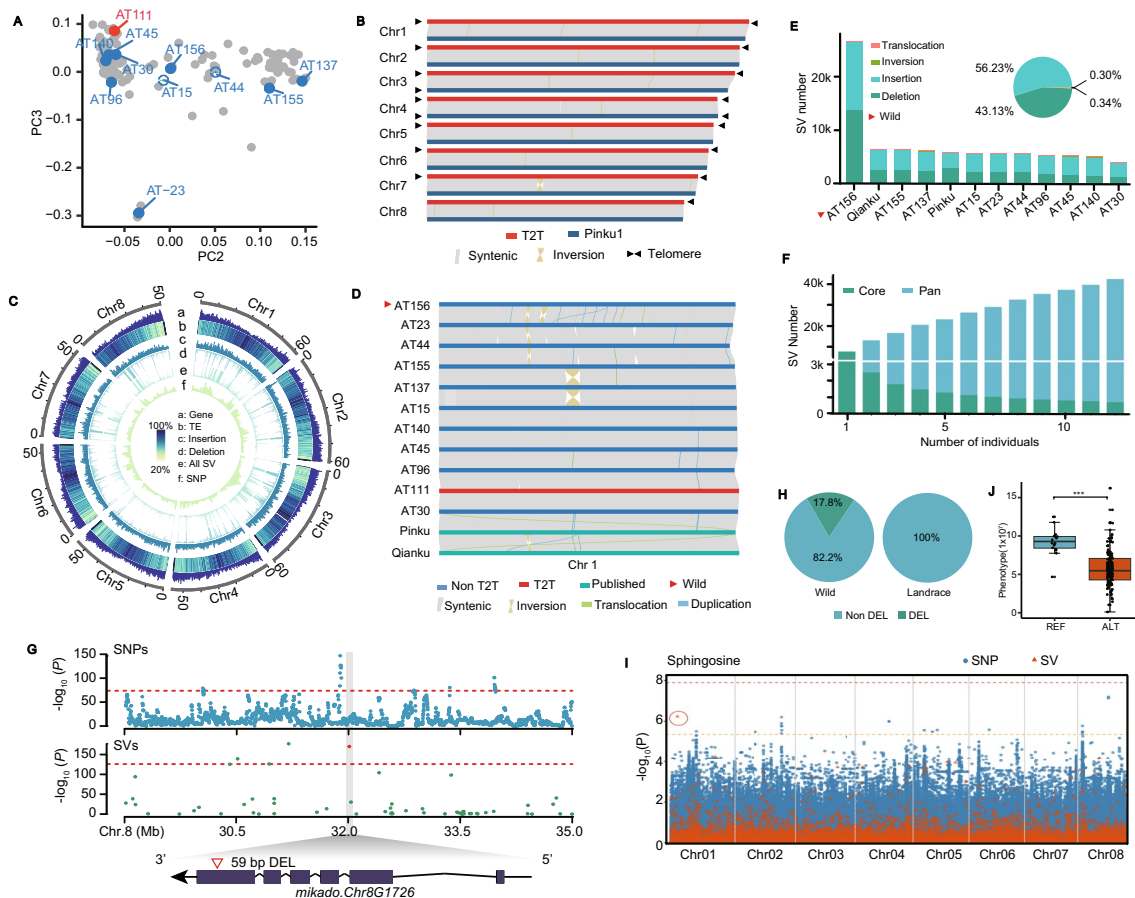


FIGURE 1 | Telomere-to-telomere assembly of the *Heifeng1* genome and overview of the graph-based pan-genome of Tartary buckwheat. (A) Principal component analysis of 164 resequenced accessions. (B) Collinearity analysis between the *Heifeng1*-T2T assembly and *Pinku1*-2023 assembly. (C) Distribution map of genetic variations in 12 Tartary buckwheat genomes. (a) Gene density. (b) TE density. (c) Insertion, (d) deletion, (e) non-redundant SVs and (f) SNP densities across 12 genomes. (D) Synteny alignment of chromosome 1 sequence in the *Heifeng1*-T2T assembly and 12 new assemblies. (E) Number of different types of SVs identified in the 12 newly assembled accessions. (F) Pan-SV and core-SV counts with additional genomes. (G) A domestication gene is only detected by SV-based selective analysis, and (H) its frequency distributions in wild and cultivated sub-populations. (I) GWAS of sphingosine levels based on both SVs and SNPs from the 163 newly resequenced accessions. Red and orange dotted lines indicate the Bonferroni significance thresholds for SNP- and SV-based GWAS, respectively. A candidate locus identified exclusively by SV-GWAS is highlighted with a red circle. (J) Haplotype analysis of the candidate SV locus associated with sphingosine levels.

RNA-seq-guided gene prediction using data from six tissues (root, stem, leaf, flower, pod and seed) enabled comprehensive gene prediction (Table S5). We integrated predictions from multiple approaches using EvidenceModeler, yielding 27 341 high-confidence protein-coding genes and 4838 non-coding RNA genes (Table S4).

To establish a graph-based pangenome capturing *F. tataricum*'s genetic diversity, we resequenced 163 accessions (18 wild, 145 cultivated) spanning 13 Chinese provinces (Figure 1A; Figure S1 and Table S1). Based on the SNP-based phylogeny and geographic distribution, we then selected 10 phylogenetically representative accessions (1 wild, 9 cultivars) for chromosome-level assembly (Figure S8). We performed genome assembly for each of the 10 accessions using a combination of ONT long-read and Illumina short-reads (Table S6). The resulting contig assemblies achieved N50 values of 10.82–55.21 Mb (Table S7 and Figure S9). The genome sizes of these assemblies ranged from 447.2 to 453.9 Mb, with BUSCO completeness scores between 98.1% and 99.3%, in which the most complete assembly

contained only five gaps (Table S7). To further represent the major and typical genetic diversity of Tartary buckwheat, we incorporated two published genomes, *Pinku1*-2023 (He et al. 2023) and *Qianku3*-2023 (Lin et al. 2023), alongside the 10 newly assembled genomes. These highly contiguous genome assemblies enabled precise SV detection through collinear block alignment (Figure 1C,D).

Aligning the 12 genomes to the *Heifeng1*-T2T reference genome, we found that wild accessions exhibited significantly higher SV prevalence than cultivated accessions (26 701 vs. 5561 SVs/sample, $p < 1 \times 10^{-5}$; Figure 1E and Table S8). Size distribution analysis revealed 76% of PAVs were < 2 kb, and 73.37% of PAVs were driven by transposable elements (TEs) (Figure S10). We identified 166 SV hotspots showing non-random chromosomal distribution (Figure S11A). Gene ontology analysis of SV hotspot regions identified significant enrichment ($p < 0.05$, Fisher's exact test) for "ADP binding" (GO:0043531), "terpene synthase activity" (GO:0010333) and "lyase activity" (GO:0016829) (Figure S11B). Notably, one such SV hotspot on chromosome 6

contained a cluster of six nucleotide-binding site (NBS) disease resistance genes (Figure S11C). All SVs were merged to generate 87 876 non-redundant SVs, comprising 49 410 insertions, 37 900 deletions, 266 translocations, and 300 inversions, with presence/absence variations (PAVs) accounting for ~99% of the total SVs (Figures 1E and 1F). Furthermore, we have constructed the first graph pan-genome for Tartary buckwheat based on the T2T assembly and newly identified SVs.

Leveraging the graph-based pangenome, we investigated SV contributions to Tartary buckwheat domestication and phenotypic variation through genotyping 16 044 high-confidence SVs across a diverse panel of 730 accessions (84 wild, 646 landraces). This panel included 163 newly sequenced accessions described in this study, as well as 567 previously published accessions (He et al. 2024; Zhang et al. 2021; Table S9). The reliability of SV calls was confirmed through both ONT long-read mapping and PCR-based experimental validation (Figures S12 and S13).

Phylogenetic reconstruction revealed divergent clustering patterns between SV- and SNP-based approaches: wild and cultivated accessions formed overlapping clusters in SV-based trees (Figure S14) contrasting with the clear separation observed in SNP-based analyses. Population differentiation analysis identified 293 SVs under strong selection (top 5%, $-\log_{10}(p) > 125.97$), of which only 9.9% (29/293) were detected by SNP-based methods, and all were located in non-coding regulatory regions (Figures S15 and S16). A representative example involves a 59-bp exonic deletion in *mikado.Chr8G1726*, showing striking frequency divergence between the wild sub-population (17.8%) and the cultivated sub-population (0%) (Figure 1G,H). Functional characterisation revealed that this gene encodes GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, known as a positive regulator of salt stress tolerance in plants. Additionally, a 104-bp deletion was discovered in the intron of *mikado.Chr7G3091* (Figure S17), which harbours a rhamnogalacturonan lyase domain, potentially playing a role in contributing to crop development and maturation.

To further evaluate the impact of SV on phenotypic variations, we conducted genome-wide association studies (GWAS) on 10 metabolic traits using both SNPs and SVs in the 163 newly sequenced accessions (Table S10). Notably, SV-based GWAS identified a significant association signal for sphingosine levels corresponding to a 235-bp insertion on chromosome 1, which was not detected in the SNP-based analysis (Figure 1I,J).

In summary, this study establishes a near T2T assembly that surpasses previous assemblies of Tartary buckwheat. Leveraging this benchmark-quality assembly, we developed the species' first graph-based pangenome, serving as an unprecedented resource for functional genomics and precision breeding. Comparative population analysis and GWAS through this pan-genome framework reveal that structural variations represent a functionally distinct and evolutionarily important source

of genetic diversity, playing a unique role in Tartary buckwheat domestication and agronomic trait improvement.

Author Contributions

D.H., P.Z. and Q.H. conceived and designed the research. W.L. and H.L. performed the bioinformatics analyses, statistical analysis, and wrote the manuscript. P.Z., D.H., and X.Z. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We would like to thank Assoc. Prof. Zhang Liu from the Center for Agricultural Genetic Resources Research, Shanxi Agricultural University for his assistance in the collection of germplasm resources. We would like to thank Prof. Aili Wei and Dr. Fan Yang from the College of Biological Sciences and Technology, Taiyuan Normal University for their assistance in experiments and helpful discussion.

Data Availability Statement

The data that support the findings of this study are openly available in National Genomics Data Center at <https://bigd.big.ac.cn/>, reference number PRJCA036216.

References

- Du, H., and C. Liang. 2019. "Assembly of Chromosome-Scale Contigs by Efficiently Resolving Repetitive Sequences With Long Reads." *Nature Communications* 10, no. 1: 5360.
- Garg, V., A. Bohra, M. Mascher, et al. 2024. "Unlocking Plant Genetics With Telomere-To-Telomere Genome Assemblies." *Nature Genetics* 56: 1788–1799.
- He, Q., D. Ma, W. Li, et al. 2023. "High-Quality *Fagopyrum esculentum* Genome Provides Insights Into the Flavonoid Accumulation Among Different Tissues and Self-Incompatibility." *Journal of Integrative Plant Biology* 65, no. 6: 1423–1441.
- He, Y., K. Zhang, Y. Shi, et al. 2024. "Genomic Insight Into the Origin, Domestication, Dispersal, Diversification and Human Selection of Tartary Buckwheat." *Genome Biology* 25, no. 1: 61.
- Lin, H., Y. Yao, P. Sun, et al. 2023. "Haplotype-Resolved Genomes of Two Buckwheat Crops Provide Insights Into Their Contrasted Rutin Concentrations and Reproductive Systems." *BMC Biology* 21, no. 1: 87.
- Zhang, K., M. He, Y. Fan, et al. 2021. "Resequencing of Global Tartary Buckwheat Accessions Reveals Multiple Domestication Events and Key Loci Associated With Agronomic Traits." *Genome Biology* 22: 1–23.
- Zhang, L., X. Li, B. Ma, et al. 2017. "The Tartary Buckwheat Genome Provides Insights Into Rutin Biosynthesis and Abiotic Stress Tolerance." *Molecular Plant* 10, no. 9: 1224–1237.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1–S3:** pbi70333-sup-0001-Da-taS1-S3.zip