



Brief Communication

Chromosome-scale haplotype-resolved genome assembly of the autotetraploid alfalfa cultivar Bolivia

Hongkui Zhang^{1,2,3,†} , Lan Zhou^{3,4,†}, Hong Zhao^{5,†}, Jiayan Liang^{3,5}, Yongle Liu^{3,5}, Chen Wang^{1,2,3}, Sijie Sun^{1,2,3}, Lizhen Song⁴, Yu'e Zhang⁵, Youfa Cheng^{3,4,6,7,*} and Yongbiao Xue^{1,2,3,5,*} 

¹National Genomics Data Center, China National Center for Bioinformation, Beijing, China

²Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

³University of Chinese Academy of Sciences, Beijing, China

⁴State Key Laboratory of Forage Breeding-by-Design and Utilization, Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing, China

⁵Laboratory of Advanced Breeding Technology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

⁶China National Botanical Garden, Beijing, China

⁷Academician Workstation of Agricultural High-tech Industrial Area of the Yellow River Delta, National Center of Technology Innovation for Comprehensive Utilization of Saline-Alkali Land, Dongying, China

Received 3 May 2025;

revised 19 June 2025;

accepted 29 June 2025.

*Correspondence (Tel +86 10 64801181; fax +86 10 64801292; email ybxue@genetics.ac.cn (Y.X.); Tel +86 10 62836674; fax +86 10 63836691; email yfcheng@ibcas.ac.cn (Y.C.))

[†]These authors contributed equally to this work.

Keywords: alfalfa, Bolivia, genome assembly.

Alfalfa (*Medicago sativa* L.) is a perennial, self-incompatible species in the Fabaceae family that occurs as both diploid ($2n = 2x = 16$) and autotetraploid ($2n = 4x = 32$) forms. Known for its high protein content, alfalfa is often referred to as the 'Queen of Forage' and is cultivated in over 80 countries worldwide (Radović *et al.*, 2009). Recent advances in sequencing technologies have facilitated the assembly of several genomes for the *M. sativa* complex, providing essential resources for molecular breeding and evolutionary studies (Chen *et al.*, 2020; Li *et al.*, 2020; Li *et al.*, 2025; Long *et al.*, 2022; Shen *et al.*, 2020; Shi *et al.*, 2024). More recently, the construction of a pan-genome from 24 diverse accessions has further enriched these genomic resources (He *et al.*, 2025). Here, we report a haplotype-resolved genome assembly of *M. sativa* ssp. *sativa* cv. Bolivia (hereafter referred to as Bolivia), sampled from an elevation of 3380 m in Bolivia (Figure S1).

We utilized high-coverage PacBio high-fidelity reads and Oxford Nanopore Technologies ultra-long reads for de novo assembly (Figure S2; Table S1). The genome assembly contained 11 803 contigs spanning 3.23 Gb, with a contig N50 of 3.25 Mb. After removing plasmid-derived sequences, Hi-C data were used to scaffold the contigs into 32 chromosomal groups. Ultimately, 3.03 Gb of contigs (96.1%) were anchored to 32 chromosomes (Figure 1a). Assembly quality was assessed by calculating the quality value (QV) using Illumina reads, yielding a QV of 51.5. K-mer completeness ranged from 54.51% to 58.29% for individual haplotypes, and 98.16% for the combined genome, indicating substantial haplotypic divergence. Despite this divergence, the Hi-C interaction matrix demonstrated robust chromosomal scaffolding (Figure S3). The overall and properly mapping rates

of Illumina reads were 99.9% and 97.9%, with uniform coverage across the genome (Figure S4). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis indicated high completeness across all haplotypes (Table S2). Phasing accuracy was assessed using both analytical and experimental approaches, revealing a low error rate of 0.73% (Figure 1b), with all PCR-validated SNPs accurately represented across the four haplotypes (Figure S5). Transposable elements, predominantly long terminal repeat (LTR) retrotransposons, comprised approximately 67.8% of the genome and the resulting LTR Assembly Index (LAI) values ranged from 16.81 to 22.45 across haplotypes (Table S2), supporting a reference-quality assembly. A total of 138 913 protein-coding genes were predicted, with BUSCO completeness reaching 98.3%. High collinearity in synteny analysis with other *Medicago* genomes further validated the assembly's structural accuracy and completeness (Figure 1c; Figure S6).

To investigate inter-haplotype variation, genes were categorized into four groups based on their synteny conservation across the haplotypes (Figure 1d). This classification reflects the degree of haplotype conservation rather than mere presence/absence variation. A similar distribution was observed in the Xinjiangdaye and Zhongmu-4 assemblies, although the proportion of Group I genes showed slight variation, likely reflecting differences in assembly quality or intrinsic genomic diversity. Functional enrichment analysis revealed that Group IV genes were consistently enriched in cytoskeletal motor activity across all three assemblies (Figure S7). We also examined the expanded and contracted gene families in 15 genomes and performed enrichment analysis (Figure 1e–g). In Bolivia, expanded gene families were primarily enriched in terms suggesting responses to oxidative stress (manganese ion binding and lipid oxidation) and DNA repair mechanisms (double-stranded DNA binding). In contrast, the enrichment profile of the high-yield variety Zhongmu-4 pointed towards a highly efficient energy production and photosynthetic system (ATP biosynthesis and Photosystem I). A parallel analysis of genome-specific genes reflected the distinct adaptive pressures on each variety (Figure 1h; Figure S8). Bolivia showed a significant enrichment in genes related to carbohydrate metabolism and energy mobilization (beta-amylase activity and polysaccharide catabolic process), suggesting an adaptive strategy to ensure rapid energy supply under the hypoxic and cold

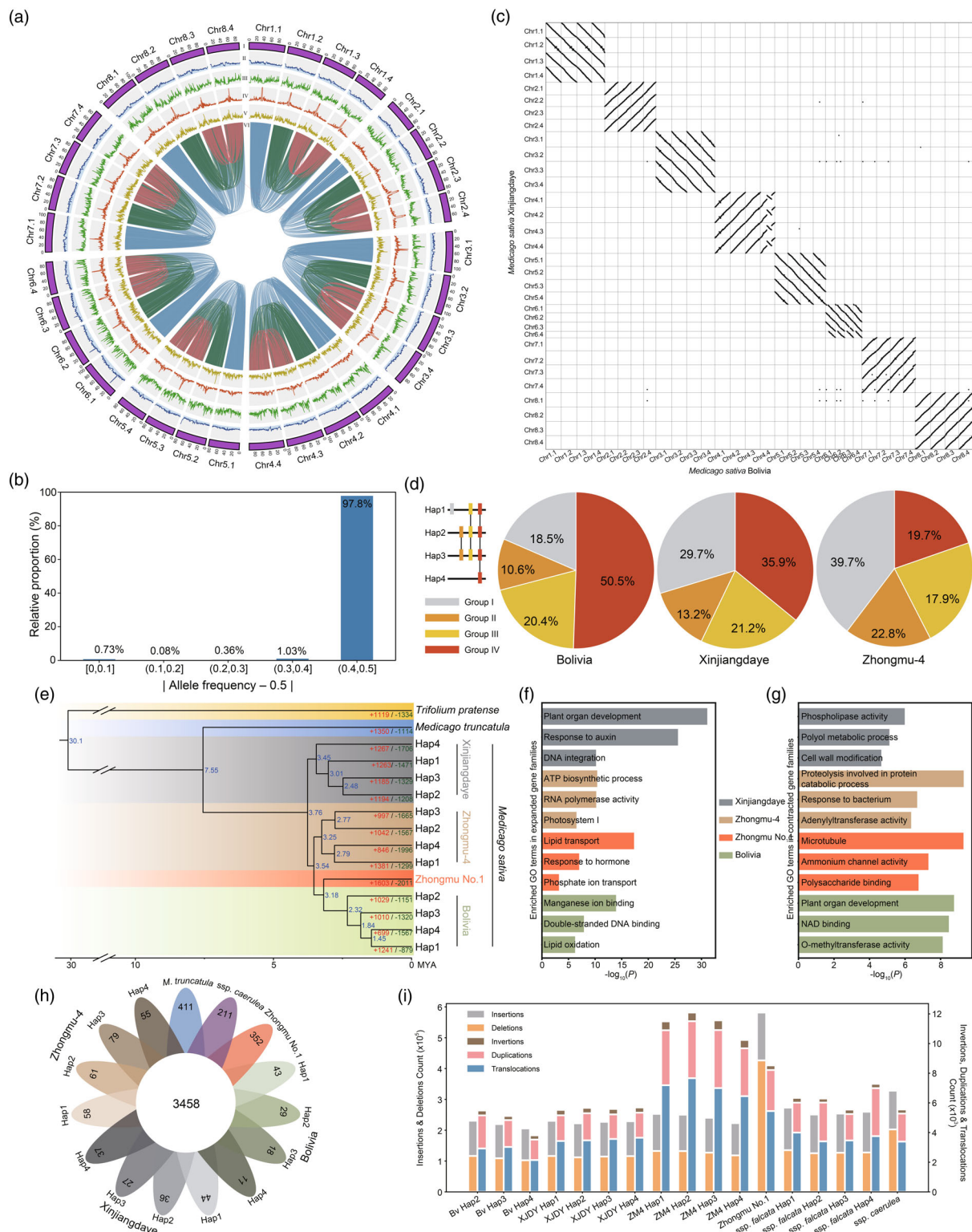


Figure 1 Haplotype-resolved genome assembly and comparative genomics of *Medicago sativa* Bolivia. (a) Circos plot of the phased genome. Tracks from outermost to innermost: (I) chromosomes, (II) GC content, (III) gene density, (IV) Gypsy density, (V) Copia density and (VI) intra-genomic syntenic blocks. (b) Phasing quality assessment. Allele frequencies near 0.5 indicate potential haplotype merging. Absolute deviations from 0.5 reflect phasing accuracy. (c) Genome synteny between Bolivia and Xinjiangdaye. (d) Gene classification based on synteny conservation. The pie chart shows the proportion of genes in four groups: Group I (one haplotype), Group II (syntenic in two haplotypes), Group III (three haplotypes) and Group IV (four haplotypes). (e) Phylogenetic tree of 15 Trifolieae genomes, showing gene family expansions and contractions. Blue numbers at nodes indicate estimated divergence times (Mya). Expansions and contractions are indicated by red and green numbers, respectively. (f) Enriched Gene Ontology (GO) terms among expanded gene families. (g) Enriched GO terms among contracted gene families. (h) Venn diagram of shared and unique gene families among 15 *Medicago* genomes. (i) Counts of different structural variant types detected in *M. sativa* genomes, using the Bolivia Hap1 as reference.

stress conditions typical of its environment. In contrast, Zhongmu-4, a high-yield and salt-tolerant cultivar, was enriched in genes for energy production (ATP biosynthetic process) and ion transport (cation channel complex). Notably, this functional signature is highly consistent with the enrichment profile of its expanded gene families, reinforcing the genetic basis for its elite traits. Finally, using the first haplotype as a reference, we identified structural variations (SVs) relative to other genomes. This analysis revealed distinct patterns of structural variation: Xinjiangdaye exhibited a higher number of translocations, Zhongmu No.1 showed more deletions and Zhongmu-4 presented an excess of translocations, duplications and inversions, highlighting the dynamic genomic rearrangements that differentiate these cultivars (Figure 11).

To address software incompatibilities and algorithmic constraints, we generated a chimeric monoploid assembly with annotated telomeres and predicted centromeres (Figures S2, S9; Table S2).

In summary, we have generated a high-quality, haplotype-resolved genome assembly of *M. sativa* ssp. *sativa* cv. Bolivia, revealing extensive haplotype-specific variation and genomic features relevant to high-altitude adaptation. While phasing remains imperfect and haplotype lengths vary, future improvements in sequencing and assembly algorithms are expected to enable gapless, telomere-to-telomere haplotype assemblies.

Acknowledgements

We thank Professor Xiqing Ma of China Agricultural University for providing the seeds of alfalfa Bolivia. This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA26030302) and National Natural Science Foundation of China (31921005).

Conflict of interest

The authors declare no competing interests.

Author contributions

Y.X. and Y.C. designed the research. L.Z., L.S. and Y.C. collected and provided plant materials. Y.Z., L.Z. and H.Z. contributed to genome sequencing. H.K.Z., L.Z., H.Z., C.W. and S.S. assembled and annotated the genomes. H.K.Z., J.L. and Y.L. performed the experiments. H.K.Z., L.Z., H.Z., Y.C. and Y.X. contributed to writing and revising the manuscript.

Data availability statement

The raw sequencing data, genome assemblies and annotations have been deposited in the National Genomics Data Center (NGDC) under the accession number PRJCA033031.

References

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F. *et al.* (2020) Allele-aware chromosome-level genome assembly and efficient

transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat. Commun.* **11**, 2494.

He, F., Chen, S., Zhang, Y., Chai, K., Zhang, Q., Kong, W., Qu, S. *et al.* (2025) Pan-genomic analysis highlights genes associated with agronomic traits and enhances genomics-assisted breeding in alfalfa. *Nat. Genet.* **57**, 1262–1273.

Li, A., Liu, A., Du, X., Chen, J.Y., Yin, M., Hu, H.Y., Shrestha, N. *et al.* (2020) A chromosome-scale genome assembly of a diploid alfalfa, the progenitor of autotetraploid alfalfa. *Hortic. Res.* **7**, 194.

Li, W., Li, X., Li, W., Yang, H., Guo, D., Guo, T., Meng, Y. *et al.* (2025) A haplotype-resolved genome assembly of tetraploid *Medicago sativa* ssp. *falcata*. *Sci. China Life Sci.* **68**, 1186–1189.

Long, R., Zhang, F., Zhang, Z., Li, M., Chen, L., Wang, X., Liu, W. *et al.* (2022) Genome assembly of alfalfa cultivar Zhongmu-4 and identification of SNPs associated with agronomic traits. *Genomics Proteomics Bioinform* **20**, 14–28.

Radović, J., Sokolović, D. and Marković, J. (2009) Alfalfa-most important perennial forage legume in animal husbandry. *Biotechnol. Animal Husbandry* **25**, 465–475.

Shen, C., Du, H., Chen, Z., Lu, H., Zhu, F., Chen, H., Meng, X. *et al.* (2020) The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasm provide genomic resources for alfalfa research. *Mol. Plant* **13**, 1250–1261.

Shi, K., Dong, H., Du, H., Li, Y., Zhou, L., Liang, C., Sakiroglu, M. *et al.* (2024) The chromosome-level assembly of the wild diploid alfalfa genome provides insights into the full landscape of genomic variations between cultivated and wild alfalfa. *Plant Biotechnol. J.* **22**, 1757–1772.

Supporting information

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

Figure S1 Global geographic distribution and elevation of *Medicago sativa* accessions.

Figure S2 The genome assembly workflow.

Figure S3 Hi-C contact matrix showing chromosomal interactions in the haplotype-resolved genome assembly.

Figure S4 Coverage depth of the assembled Bolivia genome.

Figure S5 Sanger sequencing results of target regions from four haplotypes spanning all eight chromosomes, with SNPs highlighted by black frames and triangles.

Figure S6 Genome synteny between Bolivia and representative *Medicago* genomes.

Figure S7 GO enrichment analysis of genes from different synteny groups in *M. sativa*.

Figure S8 GO enrichment analysis of unique gene families among different genomes.

Figure S9 Monoploid genome assembly and functional characterization of *M. sativa* Bolivia.

Table S1 Summary of sequencing data sizes for genome assembly of *M. sativa* Bolivia.

Table S2 Summary of the genome assembly features of *M. sativa* Bolivia.

Table S3 Primers used in this study.