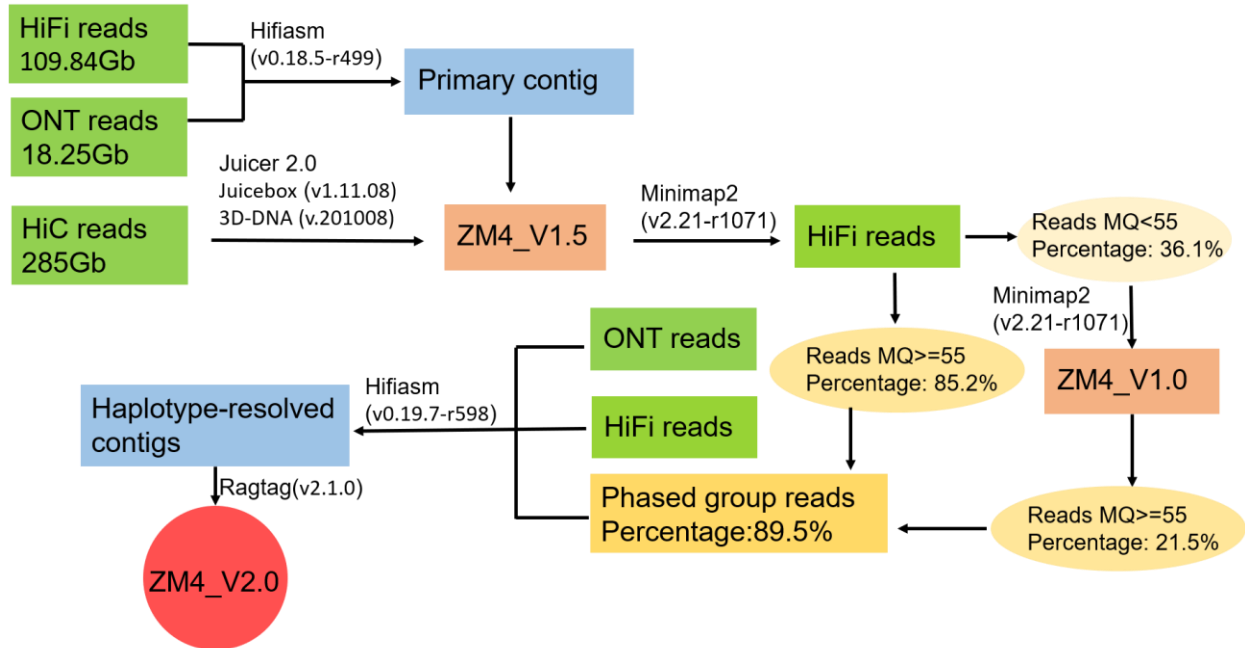
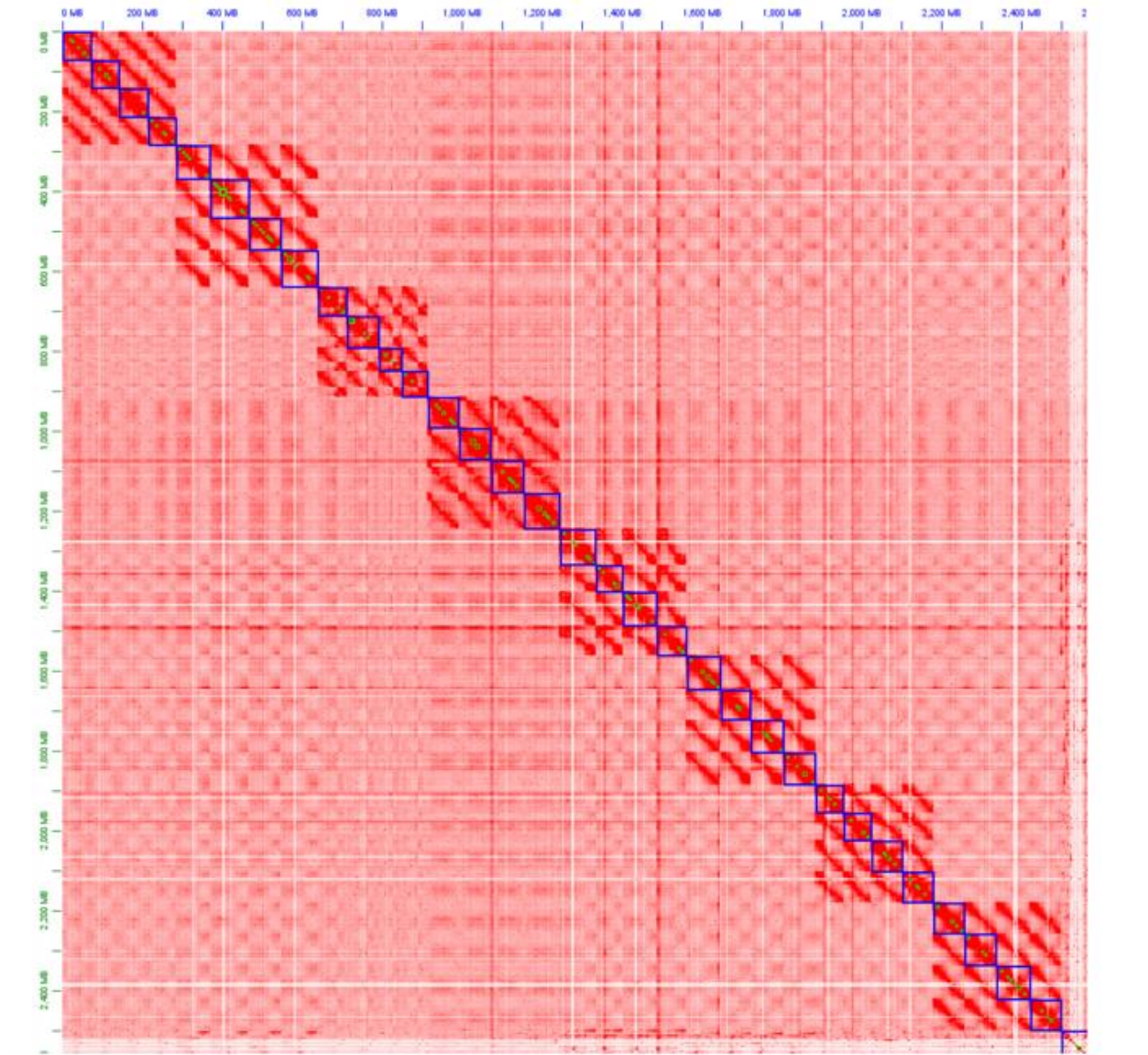


***Medicago* super-pangenome reveals adaptive advantages and evolutionary constraints in autotetraploid alfalfa**

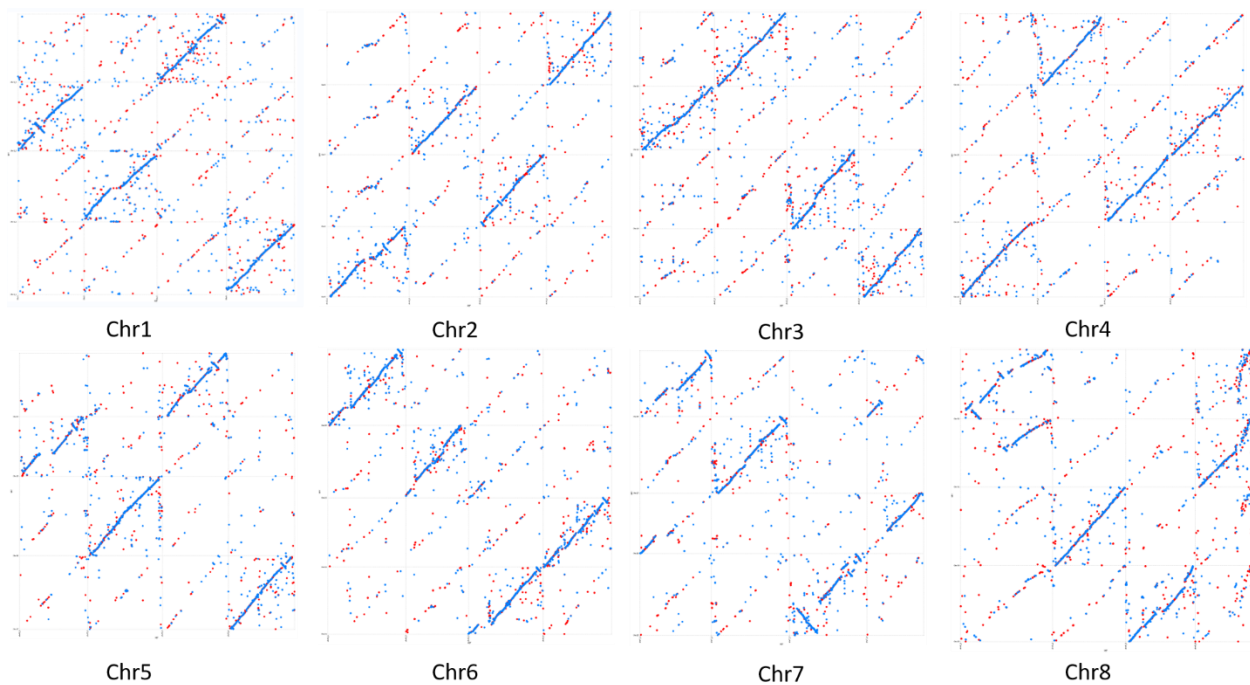
Zhang *et al.*



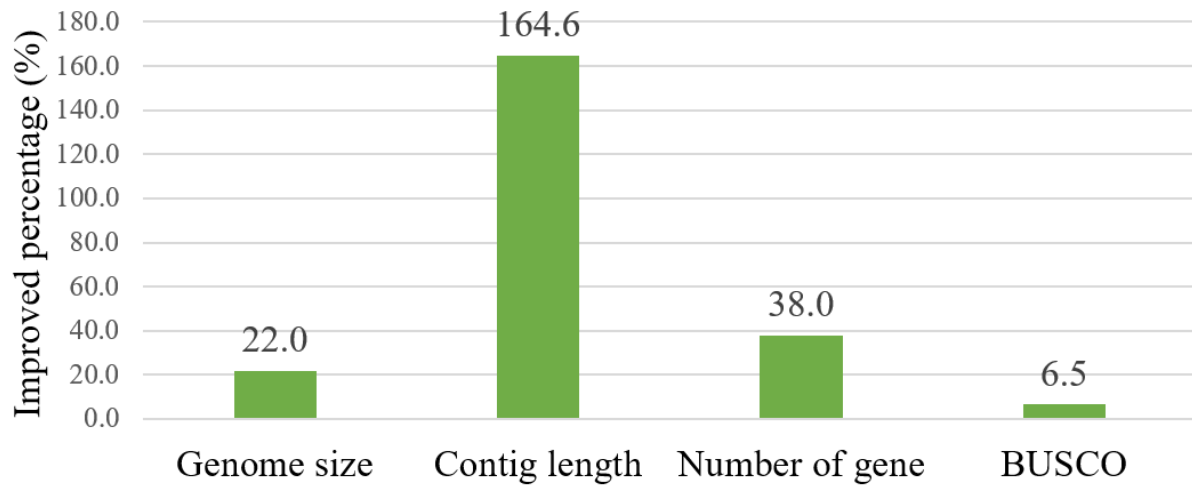
Supplementary Fig. 1. The workflow for the genome assembly of ZM4_V2.0. The software and versions used are indicated next to the corresponding steps. The ZM4_V1.0 genome is our previously published genome¹. The total number of HiFi reads was 7,352,539, and the corresponding percentages of HiFi reads with different mapping qualities (MQ) are shown. It should be noted that a single read may map to different regions of the genome and consequently receive different MQ. We utilized all mapping information for reads with $MQ \geq 55$.



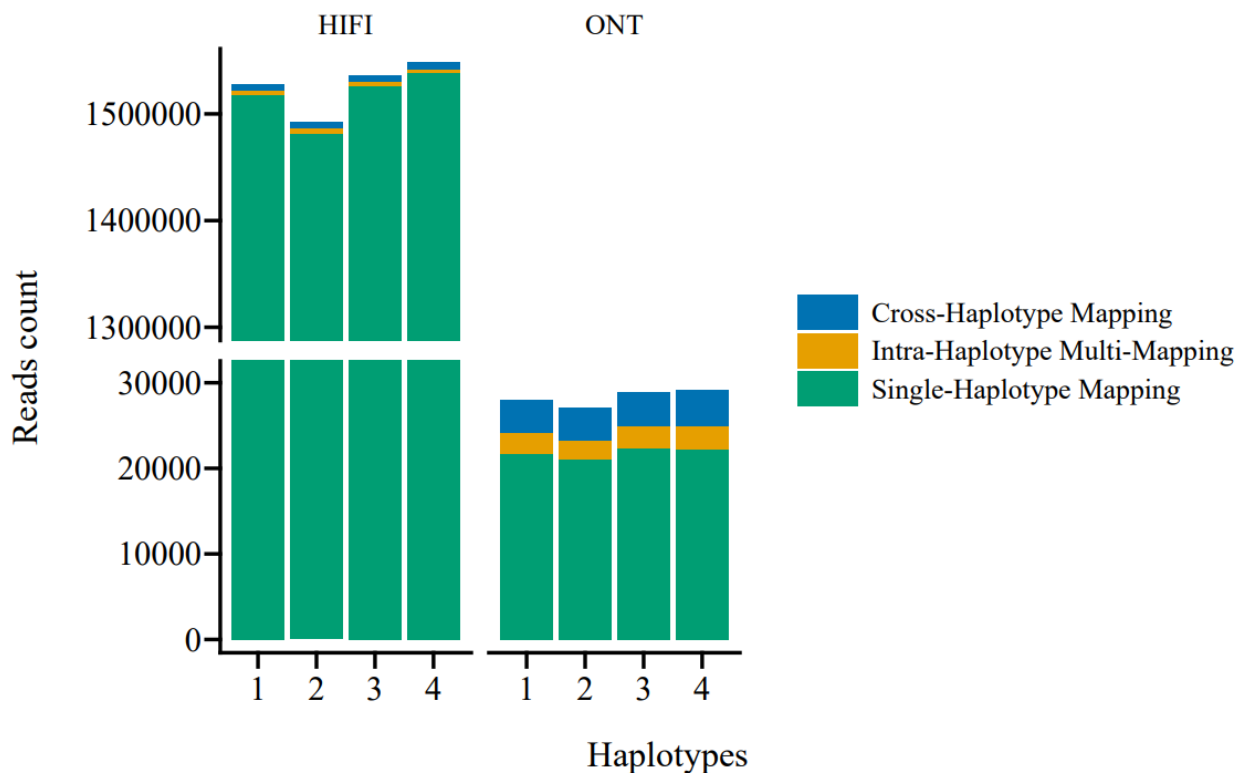
Supplementary Fig. 2. The Hi-C heatmap of ZM4_V1.5 genome. Each allelic group consists of four haplotypic chromosomes, with unclassified contigs positioned at the bottom right of the figure.



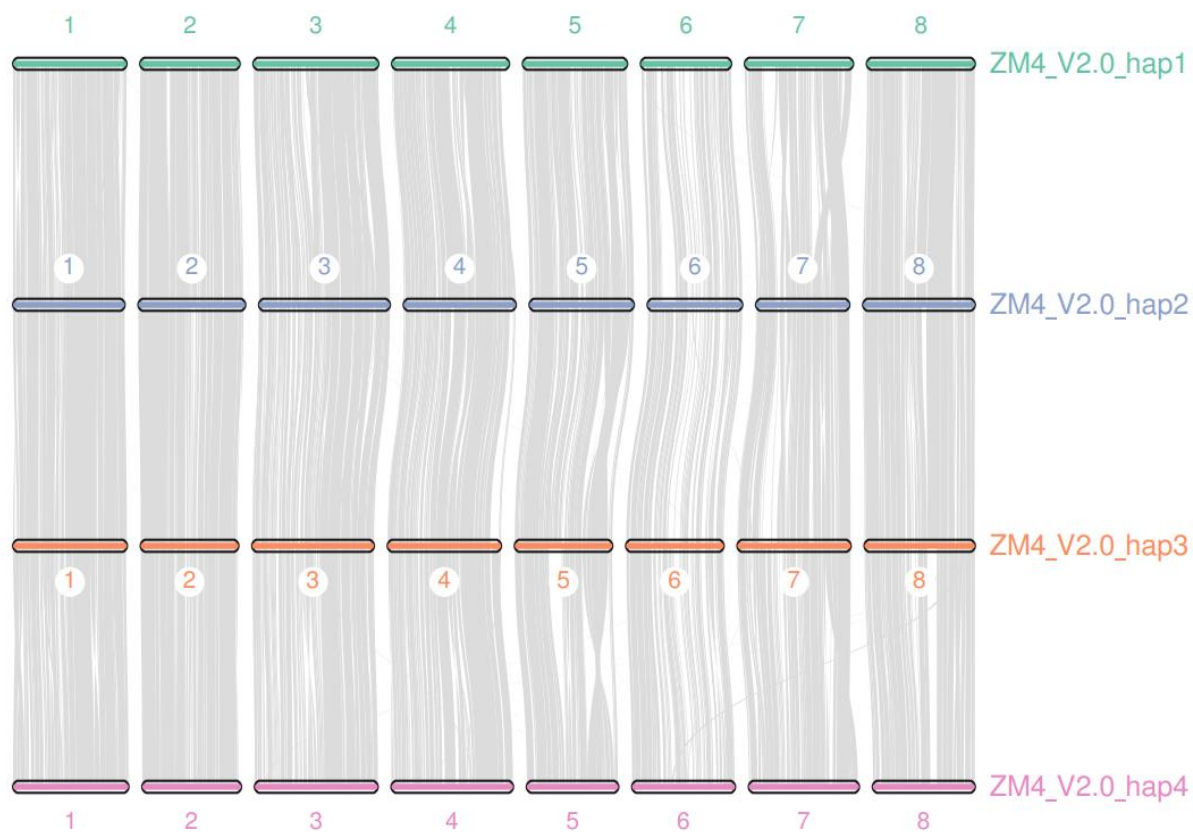
Supplementary Fig. 3. The haplotype correspondence relationships between the ZM4_V1.0 and ZM4_V2.0 genomes. The haplotype correspondence was used as the standard for phasing HiFi reads. In cases of conflicts in phasing, ZM4_V1.5 was used as the standard. The red points represent forward mapping reads, while the blue points indicate reverse mapping reads. Each point on the graph corresponds to a 4 kb segment.



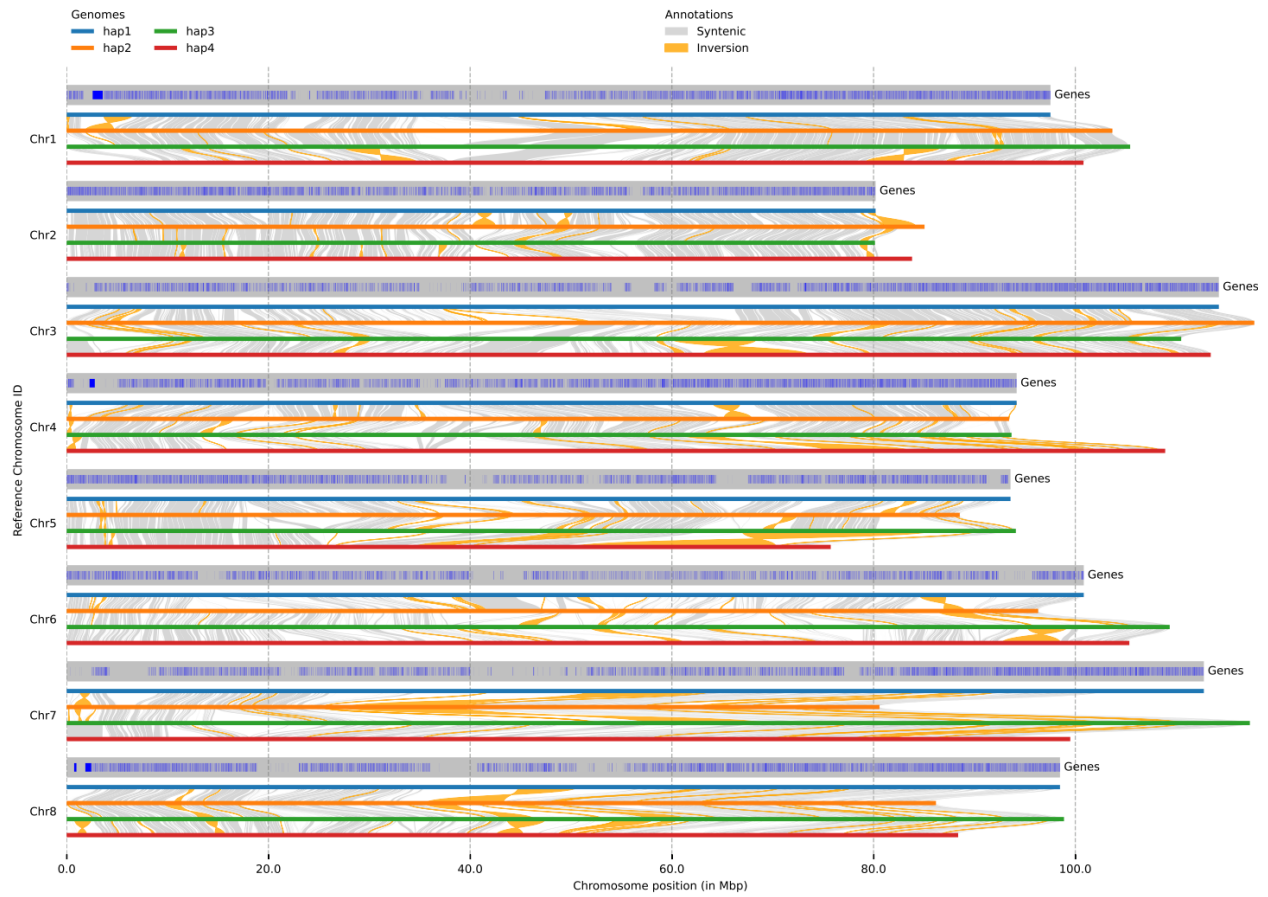
Supplementary Fig. 4. The comparison of genome quality between ZM4_V1.0 and ZM4_V2.0. Each bar represents the percentage improvement in ZM4_V2.0 relative to ZM4_V1.0. Source data are provided as a Source Data file.



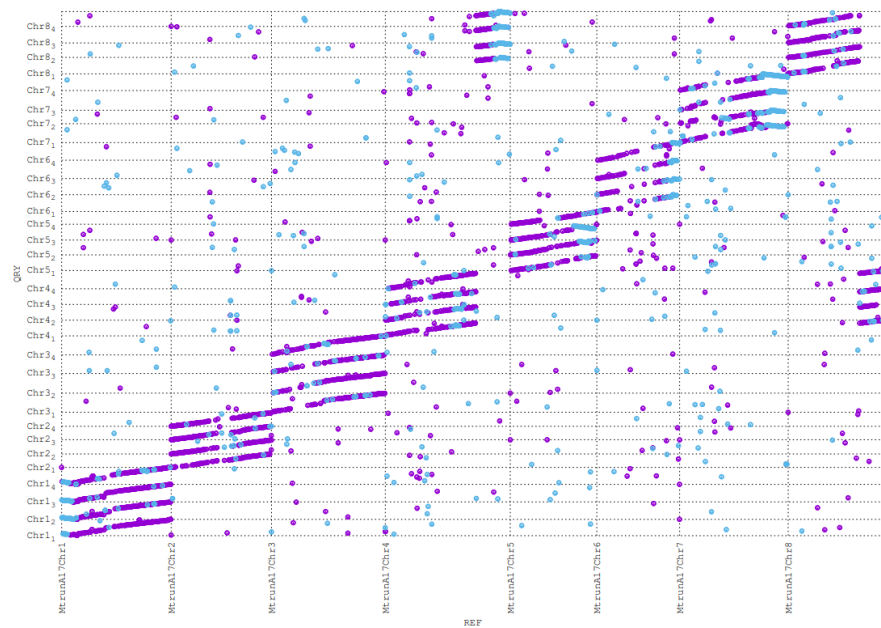
Supplementary Fig. 5. Alignment of HiFi and ONT reads to the ZM4_V2.0 reference genome. The read alignments were filtered with $\text{MAPQ} \geq 10$. Different colors are used to distinguish among three read mapping categories: (i) single-haplotype mapping, where a read maps uniquely to one haplotype; (ii) intra-haplotype multi-mapping, where a read maps to multiple locations within the same haplotype; and (iii) cross-haplotype mapping, where a read maps to different haplotypes. Overall, the final mapping rates for reads passing this quality filter were 83.02% for HiFi and 88.34% for ONT reads. Source data are provided as a Source Data file.



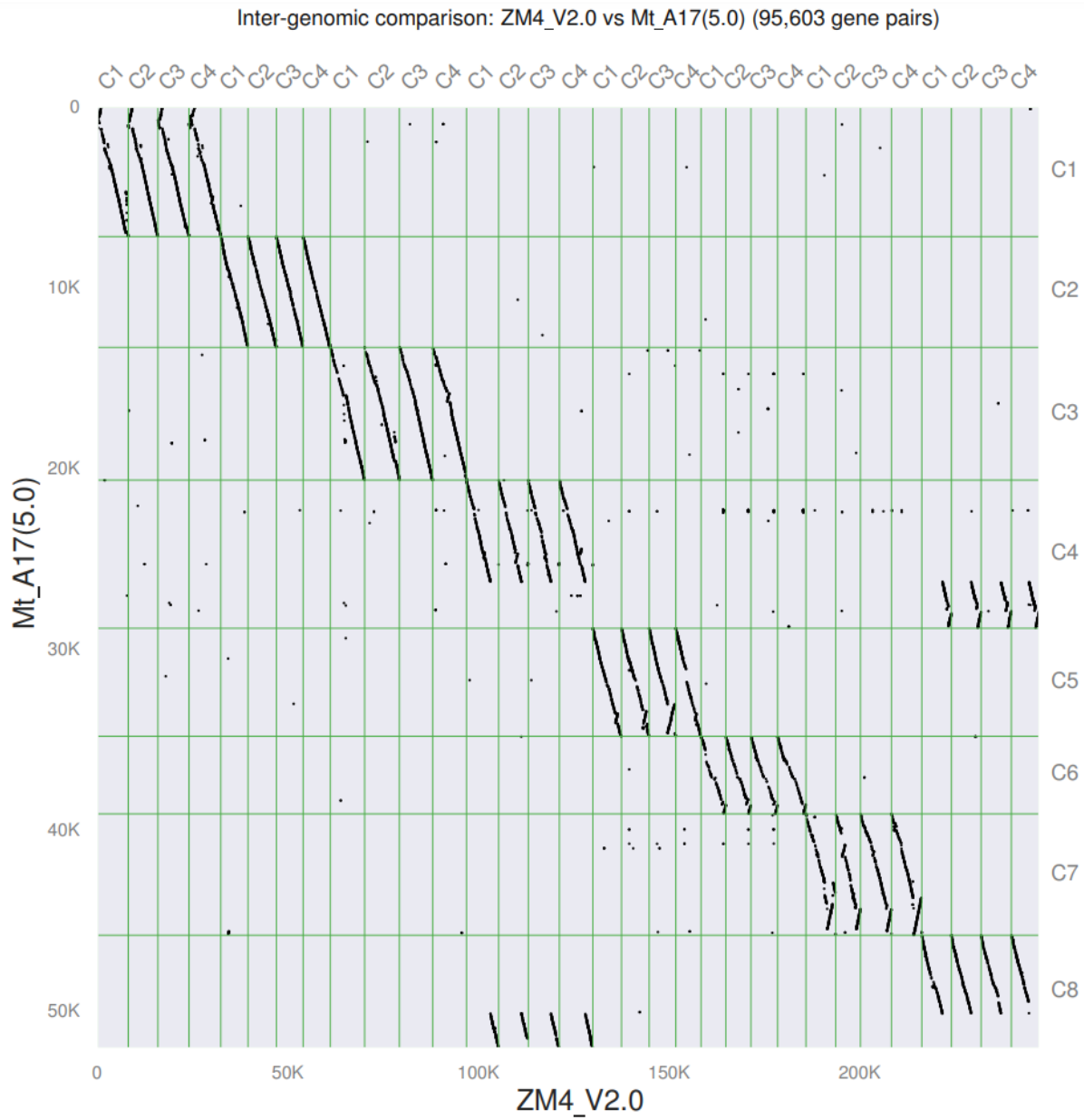
Supplementary Fig. 6. The syntenic blocks demonstrated gene synteny among the four haplotypes of the ZM4_V2.0 genome.



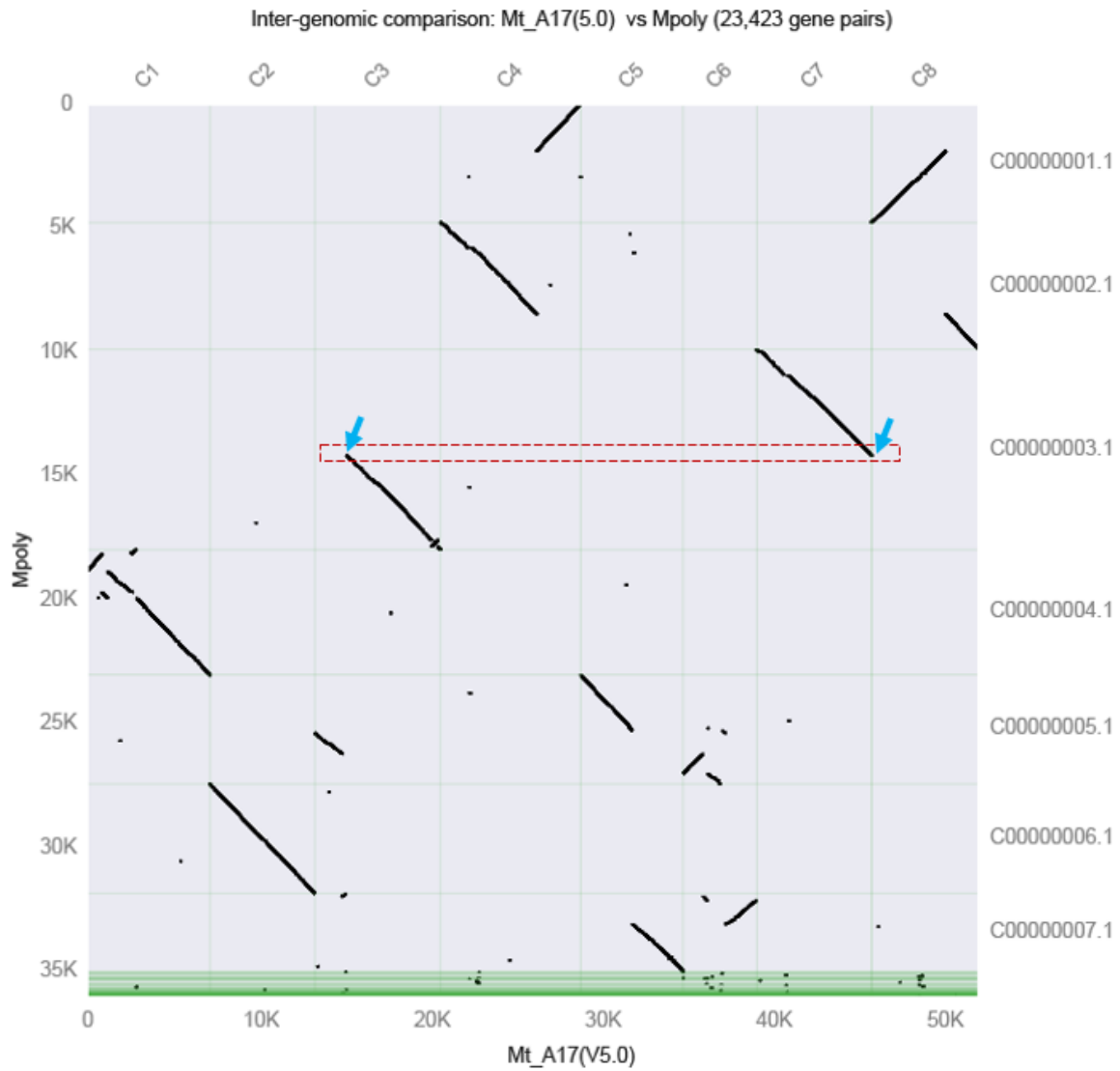
Supplementary Fig. 7. The genomic synteny and gene density throughout the alfalfa ZM4_V2.0 genome.



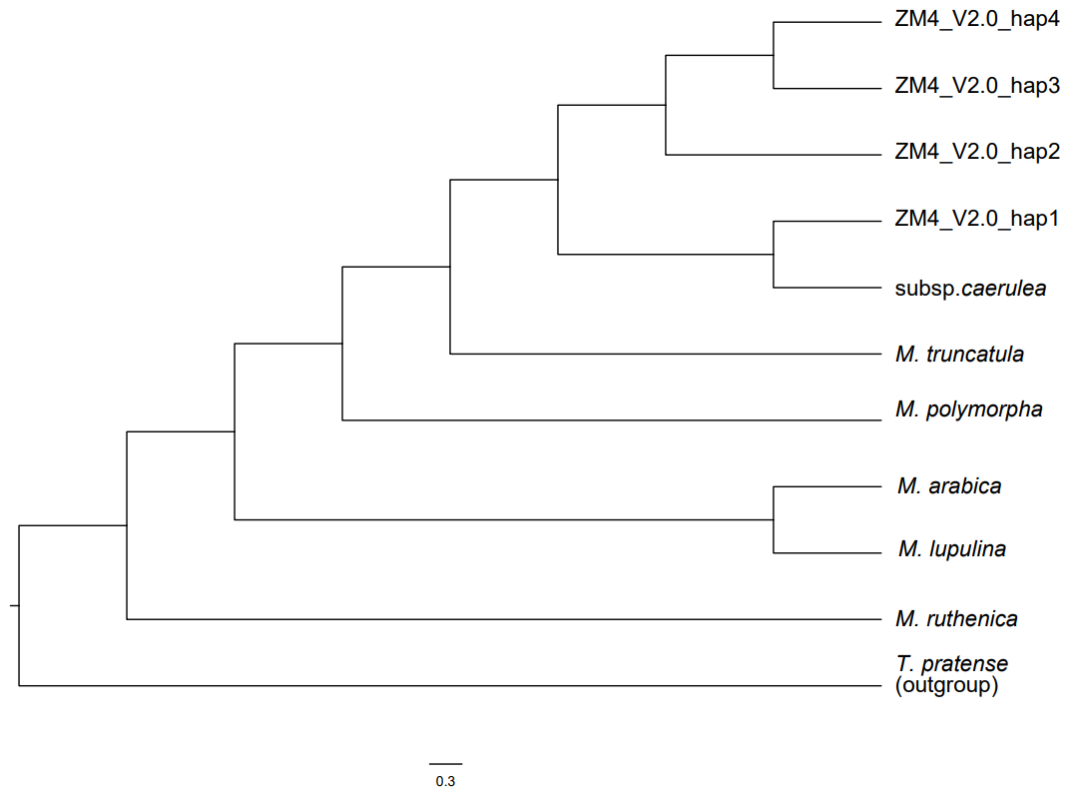
Supplementary Fig. 8. The genomic collinearity between the *M. truncatula* A17 (5.0) (REF) and ZM4_V2.0 (QRY). The purple points represent forward mapping reads, while the blue points indicate reverse mapping reads. Each point on the graph corresponds to a 5 kb segment.



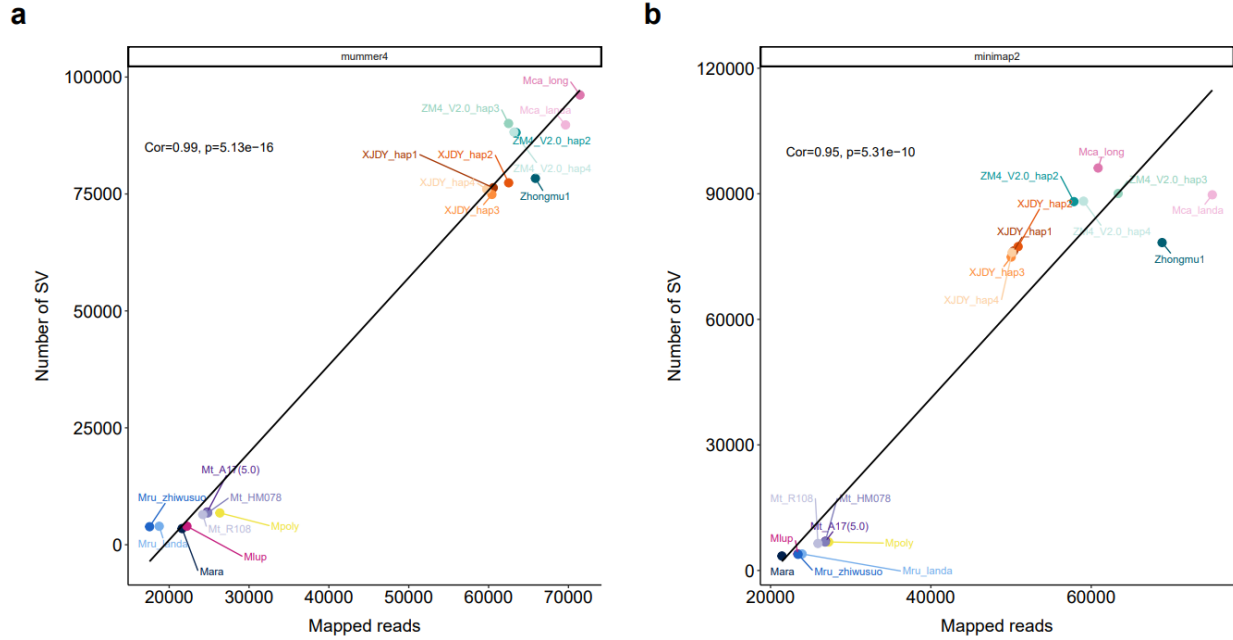
Supplementary Fig. 9. The synteny plot between the *M. truncatula* A17 (5.0) and ZM4_V2.0 genomes. Each point represents one gene pair.



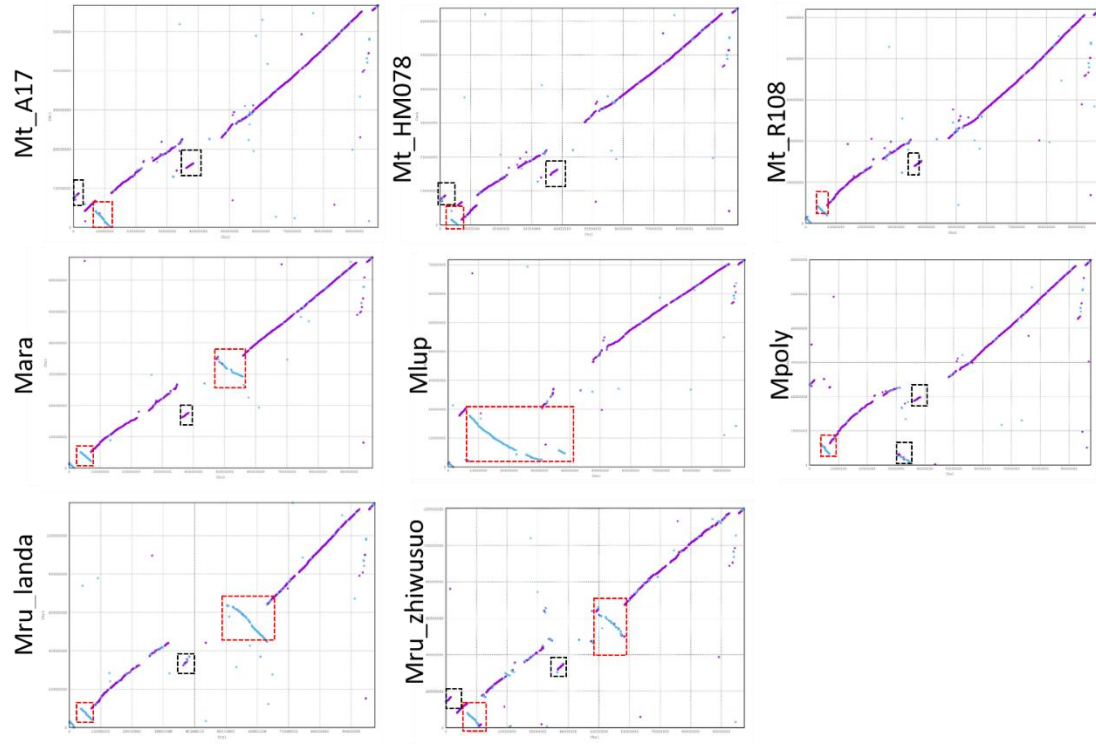
Supplementary Fig. 10. Genomic collinearity between *M. truncatula* A17 (v5.0) and *M. polymorpha*. The blue arrow indicates a breakpoint on chromosome 3, which was identified based on previously published Hi-C data². The red box highlights the surrounding region of interest.



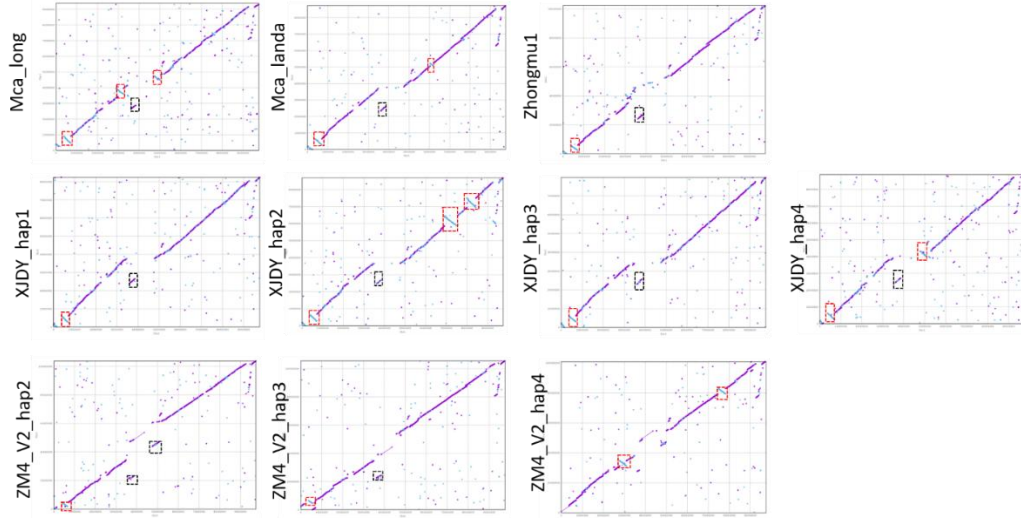
Supplementary Fig. 11. The phylogenetic tree of seven *Medicago* taxa, including alfalfa (represented by the haplotype-resolved ZM4_V2.0 genome). Source data are provided as a Source Data file.



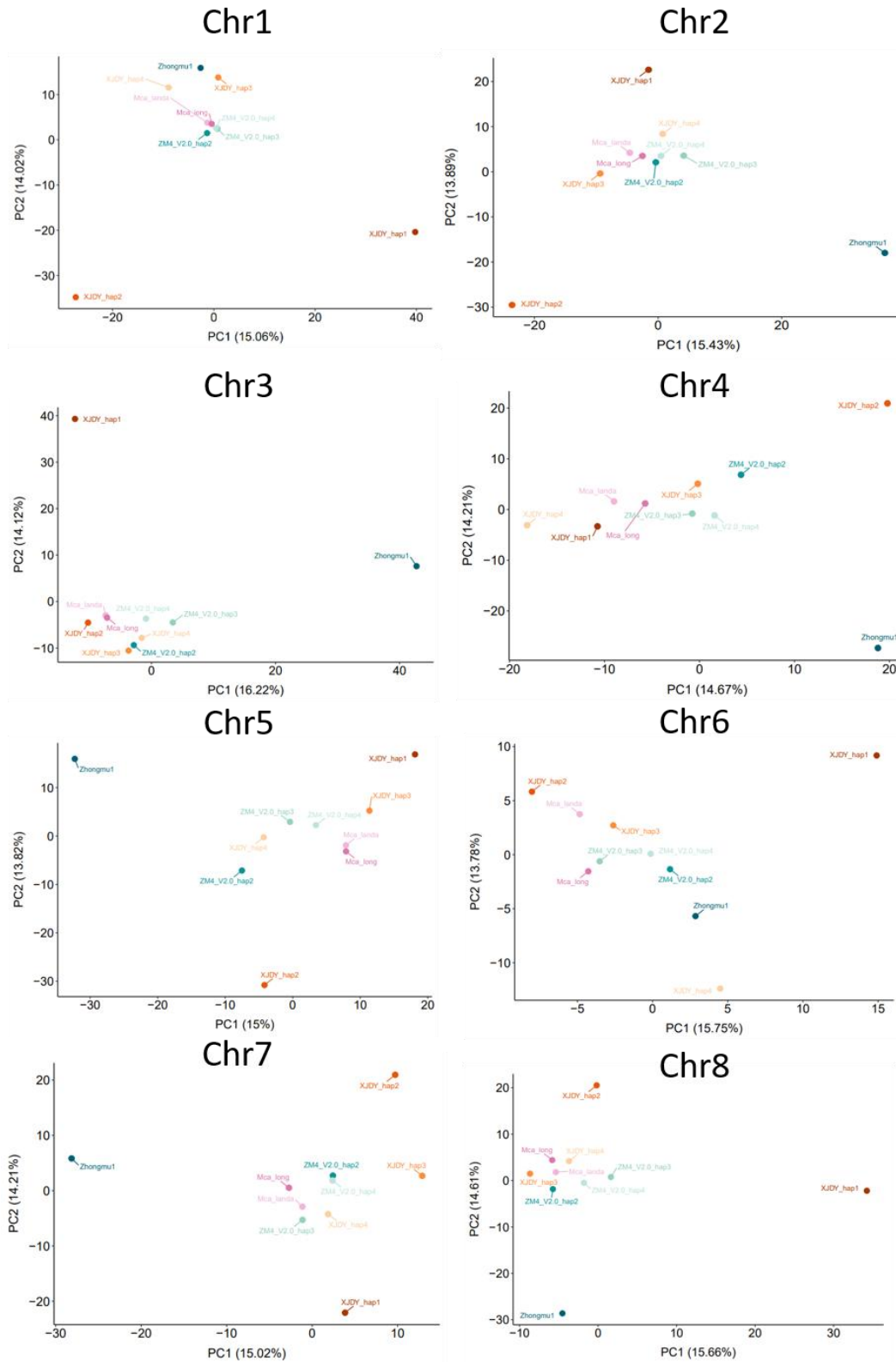
Supplementary Fig. 12. Relationship between mapped reads and the number of SVs, based on alignments generated by (a) mummer4 and (b) minimap2. The statistical relationship was assessed using a two-sided Pearson's correlation test. Source data are provided as a Source Data file.



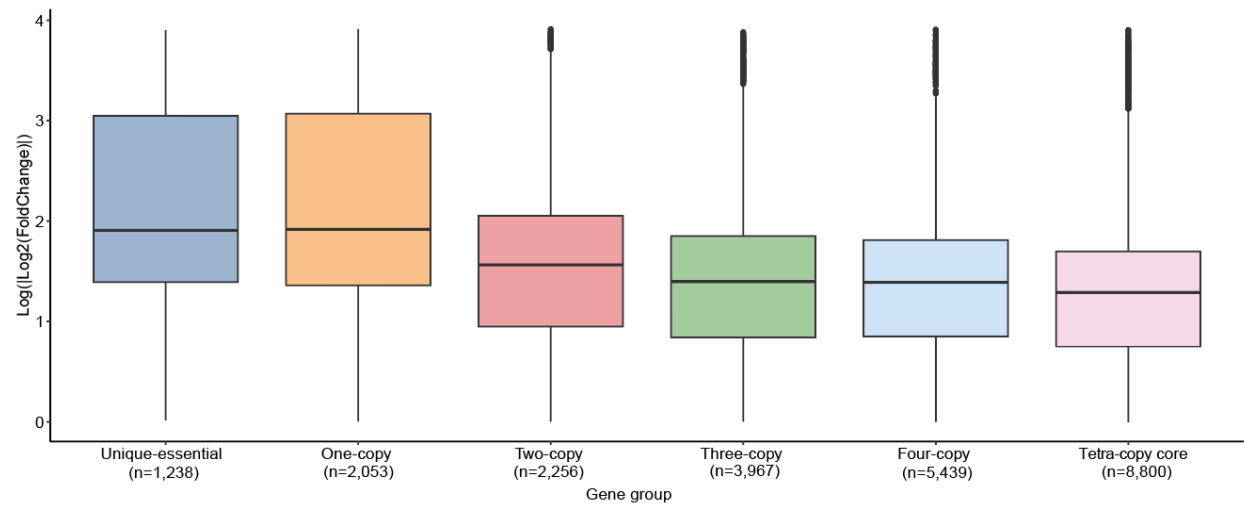
Supplementary Fig. 13. The structure variance between the ZM4_V2_hap1 (X-axis) and other *Medicago* species (Y-axis) in Chr1. Red dashed boxes represent inversions, and black dashed boxes represent translocations. The purple points represent forward mapping reads, while the blue points indicate reverse mapping reads. Each point on the graph corresponds to a 1 kb segment.



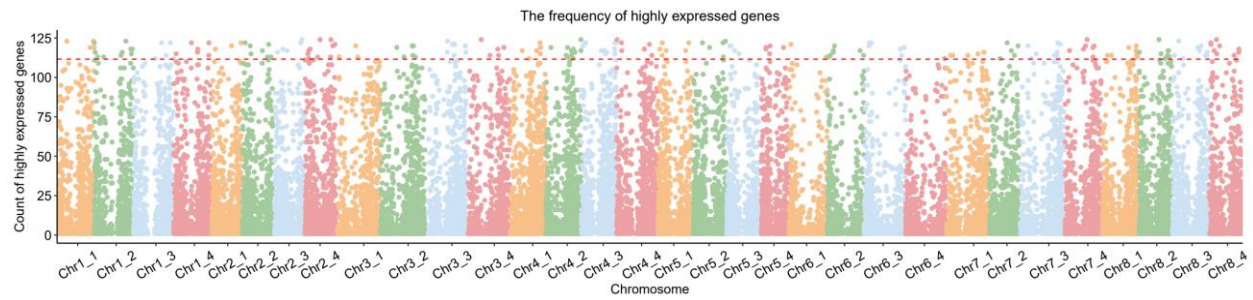
Supplementary Fig. 14. The structure variance between the ZM4_V2_hap1 (X-axis) and other *M. sativa-falcata* complex subspecies (Y-axis) in Chr1. Red dashed boxes represent inversions, and black dashed boxes represent translocations. The purple points represent forward mapping reads, while the blue points indicate reverse mapping reads. Each point on the graph corresponds to a 1 kb segment.



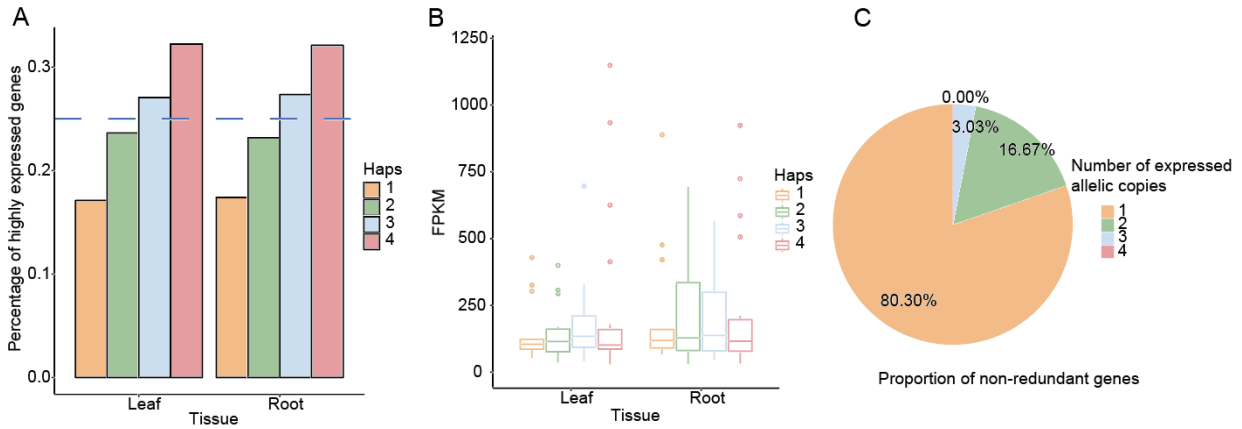
Supplementary Fig. 15. PCA of the *Medicago sativa-falcata* complex based on the haplotype-resolved genome, analyzed chromosome by chromosome.



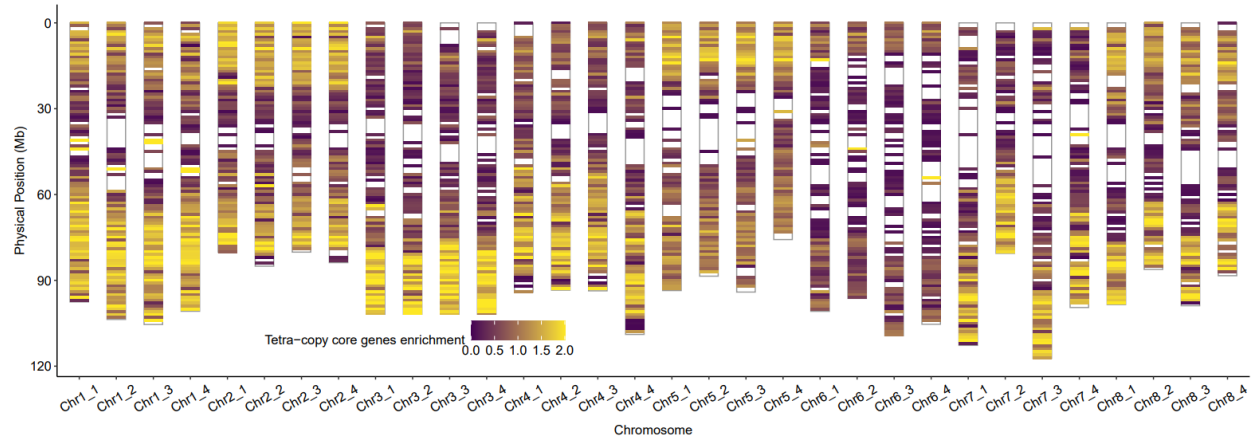
Supplementary Fig. 16. The absolute fold change values of differentially expressed genes under various stress conditions. The sample size (n) represents the number of differentially expressed genes (DEGs). Source data are provided as a Source Data file.



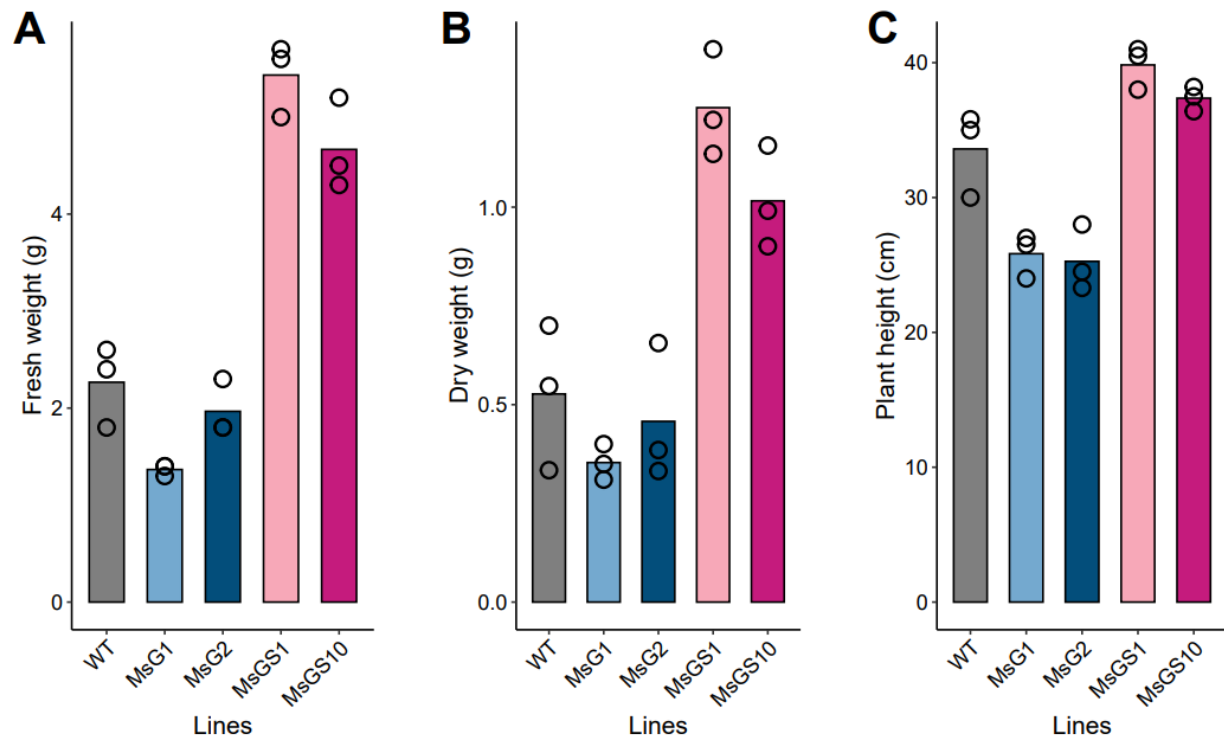
Supplementary Fig. 17. The frequency of highly expressed genes across 124 different expression conditions. The red dashed line represents the 90% threshold. Source data are provided as a Source Data file.



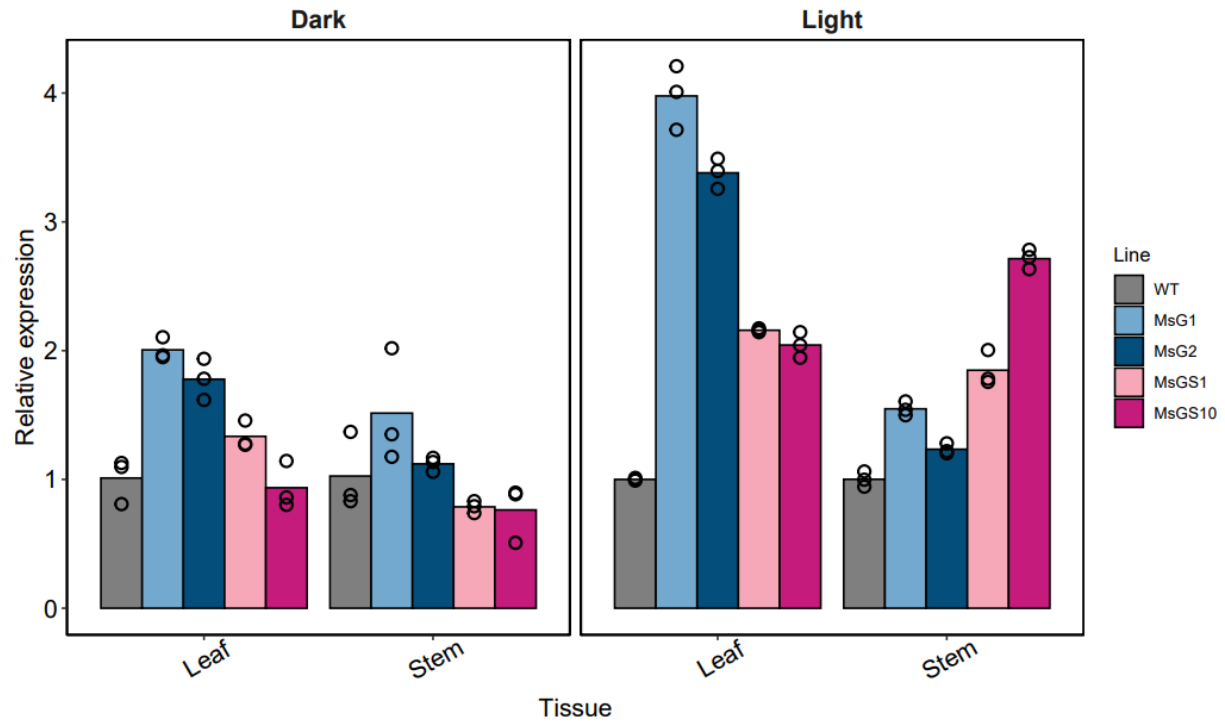
Supplementary Fig. 18. Expression patterns of consistently highly expressed Genes. Consistently highly expressed genes were defined as those ranked within the top 10% of expression across at least 90% of the experimental conditions. A, Percentage of tetra-copy core genes present in each haplotype for both leaf and root samples. The blue dish line represents 25%. B, Expression levels (FPKM values) of tetra-copy core genes in each haplotype for leaf and root samples. The sample size (n) for each boxplot is the number of genes within that category. The sample sizes were as follows: hap1 (n=14), hap2 (n=19), hap3 (n=22), and hap4 (n=26). C, Proportion of the number of expressed alleles among tetra-copy core genes. Source data are provided as a Source Data file.



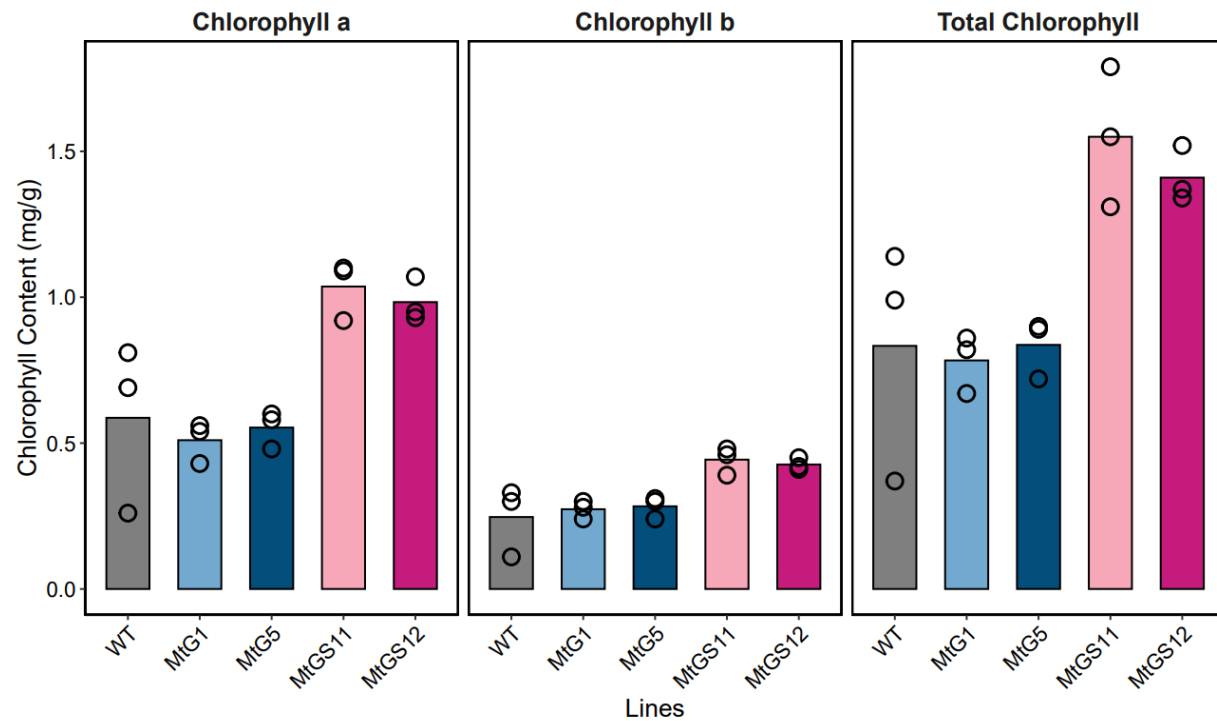
Supplementary Fig. 19. The illustration shows the distribution of tetra-copy core genes across the chromosome. The color-coded boxes represent 1Mb segment. Segments with a higher concentration of genes are shown in bright yellow, transitioning to deep purple for segments with lower gene density. A white background indicates the absence of genes. Enrichment fold exceeding 2 was set to a maximum of 2. Source data are provided as a Source Data file.



Supplementary Fig. 20. Comparison of yield-related traits between WT and transgenic *MsGDC* alfalfa. Comparison of (A) fresh weight, (B) dry weight, and (C) plant height among wild type (WT), G lines (MsG, 35S promoter), and GS lines (MsGS, ST-LS1 promoter). Each bar represents the mean of three biological replicates, with individual data points shown as dots. Source data are provided as a Source Data file.



Supplementary Fig. 21. Relative expression levels of *MsGDC* in WT and transgenic alfalfa. Each bar represents the mean of three biological replicates, with individual data points shown as dots. Source data are provided as a Source Data file.



Supplementary Fig. 22. Chlorophyll content in WT and transgenic *MsGDC* lines of *M. truncatula* under nitrogen-free treatment. Each bar represents the mean of three biological replicates, with individual data points shown as dots. Source data are provided as a Source Data file.

Supplementary Table 1. The genome assembly results for the four haplotypes of ZM4_V2.0.

Haplotype	Genome size (Mb)	Contig N50 (Mb)	BUSCO (%)[*]	Gene number	QV (base quality)	Completeness
Hap1	791.79	6.52	C:94.5, S:82.8, D:11.7	51,000	63.71	55.0%
Hap2	751.53	4.98	C:92.6, S:82.5, D:10.1	49,392	61.69	54.1%
Hap3	809.4	5.36	C:94.2, S:83.9, D:10.3	50,958	59.9	54.9%
Hap4	775.96	5.45	C:94.4, S:83.2, D:11.2	51,123	63.42	54.8%
Four_haps	3128.69	5.45	C:99.5, S:0.9, D:98.6	202,473	61.88	97.8%

^{*}C, complete BUSCOs; S, single-copy BUSCOs; D, duplicated BUSCOs.

Supplementary Table 2. The number of gene families in each pangenome category.

Categories	Count	Proportion (%)
Private	14,350	20.6
Dispensable	33,767	48.4
Softcore	8,454	12.1
Core	13,196	18.9

Supplementary Table 3. The orthologous gene families within the pangenome encompassing 13 genomes.

Genome	Short name	Number of gene families
<i>M. sativa</i> subsp. <i>sativa</i> (cv. Zhongmu No.1)	Zhongmu1	32,425
<i>M. ruthenica</i> (Institute of Botany, The Chinese Academy of Sciences)	Mru_zhiwusuo	28,097
<i>M. polymorpha</i>	Mpoly	26,997
<i>M. truncatula</i> (acc. HM078)	Mt_HM078	28,757
<i>M. truncatula</i> (acc. R108, with Hi-C)	Mt_R108	28,905
<i>M. ruthenica</i> (Lanzhou University)	Mru_landa	29,012
<i>M. arabica</i>	Mara	28,260
<i>M. caerulea</i> (Lanzhou University)	Mca_landa	33,556
<i>M. lupulina</i>	Mlup	29,319
<i>M. caerulea</i> (This study)	Mca_long	32,034
<i>M. truncatula</i> (acc. A17, v5.0)	Mt_A17(5.0)	30,352
<i>M. sativa</i> subsp. <i>sativa</i> (cv. Xinjiangdaye)	XJDY	42,646
<i>M. sativa</i> subsp. <i>sativa</i> (cv. Zhongmu No4, v2.0, This study)	ZM4_V2.0	48,151

Supplementary Table 4. The allelic copy number variation in ZM4_V2.0.

Type	Allelic copy number	Count	Proportion (%)
Non-redundant genes	Four copies	22,012	20.1
	Three copies	15,264	13.9
	Two copies	23,095	21.0
	One copy	49,344	45.0
Total genes	Four copies	79,455	39.2
	Three copies	36,330	17.9
	Two copies	37,344	18.4
	One copy	49,344	24.4

Note: The term "non-redundant genes" means that each gene is counted only once, regardless of allelic multiplicity.

Supplementary Table 5. Number of genes classified by pangenome category and allele copy number.

Allelic copy number	Categories	Non-redundant genes count	Proportion within allelic copy number group (%)
Four copies	Core	11,735	53.3
Four copies	Dispensable	2,867	13.0
Four copies	Private	662	3.0
Four copies	Softcore	5,807	26.4
Four copies	Ungrouped	941	4.3
Three copies	Core	4,950	32.4
Three copies	Dispensable	4,664	30.6
Three copies	Private	1,416	9.3
Three copies	Softcore	3,399	22.3
Three copies	Ungrouped	835	5.5
Two copies	Core	4,724	20.5
Two copies	Dispensable	9,149	39.6
Two copies	Private	3,287	14.2
Two copies	Softcore	4,438	19.2
Two copies	Ungrouped	1,497	6.5
One copy	Core	7,782	15.8
One copy	Dispensable	20,562	41.7
One copy	Private	6,610	13.4
One copy	Softcore	8,002	16.2
One copy	Ungrouped	6,388	12.9

Note: Genes not assigned by OrthoFinder to the core, softcore, dispensable, or private categories were termed 'Ungrouped'.

Supplementary Table 6. Classification of ZM4_V2.0 genes into six distinct categories.

Gene groups	Count	Proportion (%)
Tetra-copy core	43,582	21.5
Unique-essential	20,562	10.2
Four-copy	35,873	17.7
Three-copy	36,330	17.9
Two-copy	37,344	18.4
One-copy	28,782	14.2

Supplementary Table 7. The proportion of 81 highly expressed tetra-copy core genes across the four haplotypes.

Haplotype	Count	Proportion (%)
Hap1	14	17.3
Hap2	19	23.5
Hap3	22	27.2
Hap4	26	32.1

Supplementary Table 8. Primers for the functional characterization of the *MsGDC* gene.

Target region	Application	Primer name	Primer sequence (5'→3')
<i>MsGDC</i> (CDS)	Full-length gene cloning	MsGDC-F	ATGGCACTGAGGATGTGGGCTTC
		MsGDC-R	TCAATGTGCAGCATCTTCTTCCTCA
<i>ST-LSI</i> promoter	Promoter amplification	ST-LS1-F	GCAAGCTTTTGTCTATCCACTGATGTCTG
		ST-LS1-R	GCGGATCCTTGCTCTCACTACTTAGTATGA
<i>MsGDC</i> (Transcript)	Gene expression analysis	qRT-MsGDC-F	CACTGAGGATGTGGGCTTCTT
		qRT-MsGDC-R	AGCCTTCATGCTTCACCCAT

Supplementary references

1. Long, R. *et al.* Genome assembly of alfalfa cultivar zhongmu-4 and identification of SNPs associated with agronomic traits. *Genomics, Proteomics & Bioinformatics* **20**, 14-28 (2022).
2. Cui, J. *et al.* The genome of *Medicago polymorpha* provides insights into its edibility and nutritional value as a vegetable and forage legume. *Horticulture Research* **8**(2021).