

Chromosome-level genome assembly of landrace 'Huangzhou' provides insights into high cellulose accumulation in radish

Shi-Sheng Li

shishengli@hgnu.edu.cn

Huanggang Normal University

KE XU

Huanggang Normal University

Yujia Liu

Huanggang Normal University

Huan Liu

Beijing University of Agriculture

Yuanyuan Tan

Huanggang Normal University

Wei Tu

Huanggang Normal University

Hongjin Dong

Huanggang Normal University

Haiping Xin

Wuhan Botanical Garden

Changjun Ge

Huanggang Academy of Agriculture

Lei Gao

Wuhan Botanical Garden

Xiaofen Yu

Wuhan Botanical Garden

Research Article

Keywords: Genome assembly, Landrace radish, Dry matter, Cellulose, Taproot

Posted Date: June 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9939208/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on August 13th, 2026.
See the published version at <https://doi.org/10.1186/s12870-026-09729-7>.

Abstract

Radish (*Raphanus sativus* L.) is an important root vegetable crop with extensive morphological and nutritional diversity. The landrace radish, ‘Huangzhou’ (HZ) radish, origins in Huangzhou area, Hubei province of China, and characterized by a special subconical-shaped taproot and high dry matter content, making it particularly suitable for processing. However, the genetic basis underlying these agronomic traits remains largely unexplored due to the lack of a reference genome. In this study, we generated a high-quality haplotype-resolved genome assembly of HZ radish by integrating PacBio HiFi, Oxford Nanopore, and Hi-C sequencing technologies. Comparative analysis between the two haplotypes revealed substantial structural divergence and transposable elements (TEs) probably contributed to the differentiation of haplotypes. Further population analysis with eight HZ radishes from different regions, as well as another 190 radish accessions identified 21,004 unique single nucleotide polymorphism sites (SNPs) in the HZ radishes, and those related genes were enriched in response to environmental pH and aluminum ion, suggesting local environmental adaptation. Phenotypic and cytological analyses demonstrated that the high dry matter content of HZ taproots was primarily associated with increased cellulose accumulation and thicker cell walls. Transcriptome analysis across four developmental stages of taproots showed that multiple cellulose synthase-related genes were consistently upregulated in HZ compared with the low-dry-matter cultivar ‘Dachangbai’. Collectively, our research provides new genetic resources for the radish genome research, and offers new insights into the genetic mechanisms underlying dry matter accumulation and cellulose biosynthesis in radish.

1. Introduction

Radish (*Raphanus sativus* L.) is one of the most important and widely cultivated vegetables worldwide, characterized by edible taproots and high yield potential. Long-term domestication and artificial selection have generated substantial variation in root shape, size, color, texture, and flavor among radish cultivars. Radish also possesses considerable nutritional value and rich in diverse bioactive compounds, including glucosinolates, flavonoids, and polyphenols (Lee & Lee, 2006; Besma et al., 2018; Wang et al., 2022; Hur et al., 2024), which contribute to their antioxidant, anticancer, and cardiovascular protective effects (Hanlon et al., 2007; Manivannan et al., 2019; Do et al., 2021). China is the world’s largest producer of radish, with a planting area of million hectares and an annual output of 100 million tons, accounting for about 40% of the world’s cultivated area and 47% of the world’s total production (Hu et al., 2022).

Under long-term domestication and natural selection, *R. sativus* contains several varieties, such as cherry belle radish (*R. sativus* var. *radicula*), black radish (*R. sativus* var. *niger*), rat’s tail radish (*R. sativus* var. *caudatus*), oilseed radish (*R. sativus* var. *oleiformis*), and the most widely planted cultivar of radish, *R. sativus* var. *longipinnatus* (Zhang et al., 2021). Besides modern commercial cultivars, some traditional landrace radishes are still cultivated in specific local regions of China. These landrace radishes often possess specific agronomic traits, such as agronomic traits, environmental adaptations, and genetic diversity that are absent from intensively bred cultivars. Therefore, landrace radishes provide important

materials for crop improvement, and for understanding the evolutionary and domestication history of radish. Among them, ‘Huangzhou’ (HZ) radish is a traditional landrace cultivated mainly in Huanggang city of Hubei province in China. It can be easily distinguished from other varieties by subconical-shaped taproot. Due to its high content of dry matter in taproot, this landrace is mainly used for processing, such as drying and pickling (Fu et al., 2022).

Dry matter content (DMC) is an important index trait that affects the taste and nutritional quality of fresh and processed edible plants (Jiang et al., 2024). In plants, dry matter mainly consists of carbohydrates and cell-wall associated polymers, including cellulose, hemicellulose, lignin and pectin (Wang et al., 2023). The composition and structure of the cell wall strongly influence the texture and processing quality of vegetable crops (Schafer et al., 2016). Among these components, cellulose is the main component of cell walls in plants, and a key component that determines the crispness, hardness, chewiness and nutritional functions of radishes (Wang et al., 2023; Cai et al., 2026). Besides, cellulose is also an important dietary fiber which is known to be beneficial to human health (Slavin et al., 2013). Elucidating the genetic and molecular basis of cellulose accumulation is important for improving radish quality and processing characteristics. Cellulose biosynthesis in plants is mediated by cellulose synthase (CESA) complexes and organized into rosette-shaped CESA complexes (CSCs) (Wilson et al., 2021). Different CESA genes participate in the formation of primary cell walls (PCWs) and secondary cell walls (SCWs) (Taylor, 2008; Wang et al., 2020). Arabidopsis CESA1, CESA3 and CESA6-like proteins work synergistically in the synthesis of PCW cellulose, while CESA4, CESA7 and CESA8 proteins are essential for SCW cellulose synthesis (Zhang et al., 2022). Previous studies have demonstrated that transcription factors and phytohormone signaling pathways coordinately regulate cellulose biosynthesis and secondary cell wall deposition. Transcription factors, such as MYB and NAC transcription factors, act as master switches in the regulation of cell wall formation (Wang et al., 2012). In radish, however, the regulatory mechanisms underlying cellulose accumulation and high dry matter formation in taproots remain largely unclear.

Recent advances in long-read sequencing technologies have greatly accelerated the assembly of high-quality plant genomes, providing powerful tools to discover the evolution and the domestication history of crops and genetic basis of particular trait. To date, thirteen chromosome-scale genomes have been published for the *Raphanus* genus, including the cultivated radish varieties “Xinlimei”, “NAU-LB” and “C60213” (Zhang et al., 2021; Xu et al., 2023; Yang et al., 2025). Based on these genomic resource, the genetic basis underlying the bolting trait was investigated and a 647-bp insertion in the *RsVRN1* gene promoter was found to be responsible for the late-bolting phenotype (Xu et al., 2023). Another study found *RsMIPS3* gene may be also associated with bolting and flowering time (Yang et al., 2025). However, current genomic studies mainly focus on modern cultivated accessions, whereas traditional landrace radishes with unique agronomic traits remain largely underexplored. In particular, no high-quality genome resource is currently available for HZ radish, limiting the genetic dissection and utilization of its valuable traits in breeding programs.

In this study, we performed the *de novo* chromosome-scale assembly of HZ radish. The haplotype-resolved assemblies of HZ radish were generated using Pacific Biosciences (PacBio) HiFi sequencing, Oxford nanopore (ONT) sequencing and high-throughput chromatin conformation capture (Hi-C) sequencing data. Using comparative genomics and population genetic analyses, we characterized the genomic diversity and unique genetic polymorphisms of HZ radishes. Furthermore, through integrative analyses of cytological observations, cell wall composition, and transcriptomic profiles, we investigated the molecular basis underlying the high dry matter and cellulose accumulation in HZ taproots. Our study provides important genomic resources for radish research and offers new insights into the genetic regulation of cellulose biosynthesis and dry matter accumulation in HZ radishes.

2. Materials and methods

2.1 Plant materials and sample collection

The HZ radish for genome sequencing and assembly was cultivated at the experimental agricultural station of Huanggang Normal University, Hubei, China. Young leaves were collected for genomic DNA extraction and genome sequencing. Leaves, flowers and siliques were collected for RNA-seq with BGI platform.

For population resequencing analysis, seven additional HZ accessions originating from different cultivation regions were collected (Table S1). For comparative transcriptome analysis, taproots from HZ and the low-dry-matter radish cultivar 'Dachangbai' (DCB) were sampled at four developmental stages. All harvested tissues were immediately frozen in liquid nitrogen and stored at -80°C . The seed materials of HZ and DCB radish were supplied by the Biological Museum of Huanggang Normal University, with HZ and DCB designated as Huangzhou radish and Dachangbai radish, respectively.

2.2 Flow cytometry analysis

Flow cytometry analysis was used to estimate genome size and ploidy of HZ radish (Doležel et al., 2007). Fresh young leaves were obtained for preparation of a suspension of intact nuclei. The DNA-specific fluorescent dye was added, and then fluorescence intensity of each nucleus was measured.

2.3 Genome sequencing and assembly

Genomic DNA from HZ leaves was isolated by a QIAGEN kit following the standard protocol, and the quantification and purity were measured by NanoDrop and Qubit instruments. Following the recommended protocol, the genomic libraries of PacBio HiFi sequencing, ONT sequencing and BGI sequencing were prepared and sequenced on the corresponding platforms. For Hi-C sequencing, genomic DNA was treated with formaldehyde for crosslinking before extraction and library preparation. Restriction enzyme *MobI* was used and the library was then sequenced on BGI platform with PE150 mode.

For BGI reads, clean data was generated by filtering low-quality reads and removing adaptors by fastp 0.22.0 (Chen et al., 2018). The genome size and heterozygosity of the HZ radish was estimated based on k-mer approach (Binghang et al., 2013).

The PacBio HiFi reads, ONT ultra-long reads (N50 = 100 kb), and Hi-C reads (Table S1) were all used for genome assembly by Hifiasm 0.19.5-r592 (Cheng et al., 2024). The primary and the phased contigs were obtained and organelle or redundant contigs were removed. Subsequently, the draft assemblies were anchored to pseudomolecules by Juicer v1.9.9 (Durand et al., 2016) and 3D-DNA v180114 (Dudchenko et al., 2017). Manual correction and orientation adjustment were performed with Juicebox v1.11.08 (Robinson et al., 2018). The assemblies were evaluated by BUSCO v5.0 with the 'eudicots_odb10' database (Manni et al., 2021) and LTR Assembly Index (LAI) estimation (Ou et al., 2018).

2.4 Genome annotation

TEs were annotated by using the comprehensive pipeline Extensive *de novo* TE Annotator (EDTA) with default parameters (Su et al., 2021). The protein-coding genes were predicted by integrated transcriptome, *ab initio* and homologue-based predictions. For transcriptome evidence, clean RNA-seq reads of leaves, flowers and siliques were aligned to the genome assemblies using HISAT2 v2.2.1 (Kim et al., 2019) and assembled with StringTie2 v2.1.6 (Kovaka et al., 2019). Coding regions were predicted by TransDecoder v5.5.0 (<http://transdecoder.github.io>). To perform *ab initio* gene prediction, BRAKER v2.1.6 (Hoff et al., 2019) was used to generate gene structure models by integrating genomic and RNA-seq data. Protein sequences from *R. sativus* var. *longipinnatus*, *Arabidopsis thaliana*, *Solanum lycopersicum*, and *Oryza sativa* were used for homology-based prediction. All these results were integrated by MAKER v 2.31.11 (Campbell et al., 2014) to generate the final gene models. For gene functional annotation, the predicted protein sequences were blast against public protein databases, including the GO, Kyoto Encyclopedia of Gene and Genomes (KEGG), GenBank Non-redundant, UniProt and Interpro databases with DIAMOND v2.0.13.151 (Buchfink et al., 2021).

2.5 Comparative genomics and phylogenetic analysis

Whole genome alignments were performed by Minimap2 (Li H. 2018), and then structure variations and syntenic regions were identified by SyRI (Goel et al., 2019). The HiFi reads were mapped to genome assembly by Minimap2, and mapping depth and coverage were calculated using SAMtools v1.15 (Li et al., 2009). To elucidated the evolutionary history of HZ, gene families of HZ haplomes, 12 additional radish accessions (Zhang et al., 2021; Xu et al., 2023), and the outgroup *Brassica oleracea* were identified using OrthoFinder v2.5.2 (Emms et al., 2019). The single-copy orthologs were aligned by MUSCLE v3.8.31 (Edgar, 2004), and a maximum-likelihood (ML) tree was then generated by RAxML v8.2.12 (Stamatakis et al., 2014). A phylogenic tree with species divergence times was constructed by MCMC tree tool (Yang et al., 2007). The estimated divergence times for *Raphanus* genus and *B. oleracea* were calibrated according to the Time Tree database (<http://www.timetree.org/>) and published study (Zhang et al., 2021). The percentage of hemizygous genes was estimated following the approach described by Peng et al. (2025).

2.6 Genome resequencing and population genetic analysis

Genomic DNA from seven HZ accessions was extracted for resequencing library construction. Sequencing was performed on the BGI platform, generating approximately 20× coverage for each accession. After quality filtering, clean reads were mapped to the HZ primary assembly by BWA v0.7.17 (Li et al., 2009). Besides, resequencing data of 190 cultivated and wild radish accessions from two published studies (Zhang et al., 2021; Fan et al., 2020) were also filtered and mapped to the HZ primary assembly. SNP calling was performed by GATK v4.2.0 (Depristo et al., 2011). SNPs were firstly filtered with the following criteria: $QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < 12.5 || ReadPosRankSum < 8.0$ by GATK, and further filtered by VCFtools v0.1.16 (Danecek et al., 2011) for subsequent population analysis with the following parameters: -minQ 30, -maf 0.05, -max-alleles 2, -min-alleles 2, -max-missing 0.5. Principal component analysis (PCA) analysis was performed using PLINK v1.90 (Purcell et al., 2007). The phylogenetic tree was constructed by IQ-TREE with 1,000 bootstrap replicates, and displayed and annotated by online tool iTOL (Letunic et al., 2024).

2.7 Transcriptome sequencing and differential expression analysis

The taproots of the DCB and HZ radishes at four developmental stages were collected and washed, and immediately frozen in liquid nitrogen and stored at -80°C. Total RNA was then extracted using Tiangen RNAprep pure plant kit (Tiangen Biotech, China). RNA-seq libraries were constructed according to the manufacturer's instruction and sequenced on a DNBSEQ-T7 to generate 150-bp paired-end reads. Clean RNA-seq reads of each sample were aligned to the HZ primary genome assembly using HISAT2. Gene expression levels were calculated by StringTie, and read counts were further used for differentially expressed genes analysis by DESeq2 (Love et al., 2014). Differentially expressed genes (DEGs) were identified if the $|\log_2(\text{fold change})|$ was > 1 and the p_{adj} was < 0.05 . PCA analysis of sample expression pattern was performed by R package.

2.8. RT-qPCR assay

The quantitative expressions of genes involved in cellulose synthesis were analyzed by RT-qPCR. The radish RNA polymerase-II transcription factor (RP11) gene (*HZch01g037890*) was used as the internal reference gene for normalization (Xu et al., 2012). Relative gene expression level was assessed with $2^{-\Delta\Delta Ct}$ method. Primer sequences are shown in Table S2.

2.9 Detection of dry matter and principal fiber components in taproots

Dry matter content was determined using the direct drying method according to the national food safety standard GB 5009.3–2016. Briefly, fresh taproot samples were weighed before drying and reweighed after drying to constant weight. DMC was calculated as the ratio of dry weight to fresh weight. Cellulose, hemicellulose, and protopectin content were measured using spectrophotometric quantification

following acid hydrolysis. Lignin content was determined using the potassium dichromate oxidation method. Lignin oxidation quantity was calculated based on sodium thiosulphate consumption, subsequently converted to determine sample lignin content. Crude fiber content was analyzed using the acid-alkali digestion method. After accurately weighing the sample, it is first treated with boiling 1.25% sulphuric acid solution, followed by boiling 1.25% potassium hydroxide solution. The crude fiber content was calculated from the mass loss before and after digestion.

2.10 Ultrastructural analysis of taproot cells

Fresh taproot tissues were fixed with 2.5% glutaraldehyde, 2% paraformaldehyde and 0.1M phosphate buffer (pH7.4) for 4h at 4°C. Samples were dehydrated, embedded, and sectioned using an ultramicrotome (Ventana MT-X, Tucson, Arizona, USA). Semi-thin sections were stained with toluidine blue (SCRC#710412, China). Ultrathin sections were stained serially with 2% uranyl acetate, 1% lead citrate. Cellular ultrastructure was observed using a transmission electron microscope (Hitachi, H-8100, Tokyo, Japan).

3. Results

3.1 Haplotype-resolved genome assembly of Huangzhou radish

Young leaves of a Huangzhou radish individual (Hereinafter referred to as HZ, Fig. 1a) were collected, and flow cytometry analysis revealed that it was diploid with an estimated genome size of 572.13 Mb (Fig. S1). Estimates based on k-mers ($k = 17$) revealed that HZ genome had a relatively high degree of heterozygosity (1.34%). A total of 30.49 Gb (~60X) PacBio HiFi reads, 27.00 Gb (~54X) ONT ultra-long reads, and 61.10 Gb (~122X) Hi-C reads were generated for genome assembly (Table S1). All these reads were used to generate the initial assemblies by Hifiasm, resulting in a primary assembly of 489.22 Mb with a contig N50 of 43.17 Mb. Due to its high heterozygosity, we generated its haplotype-resolved assemblies named HapA and HapB, with total lengths of 462.57 Mb and 449.29 Mb, containing 69 and 60 contigs each, with contig N50 values of 12.92 Mb and 16.73 Mb, respectively (Table 1). Utilizing the Hi-C reads, each assembly was scaffolded into nine pseudomolecules (Fig. 1b-c, Fig. S2). Evaluation of the genome completeness by BUSCO analysis showed that the primary, HapA and HapB assemblies exhibited coverages of 99.1%, 97.2% and 96.6% of the embryophyta single-copy gene set, respectively. Moreover, all assemblies had LAI over 16, indicating high assembly contiguity of HZ genome assemblies.

Figure 1 Morphology and genomic features of HZ radish. GC content (i), repeat content (ii), gene density (iii) in 500-kb sliding windows, and chromosomal synteny between haplotypes (iv) are shown. Root type of HZ radish is shown.

Repeat annotation revealed that repetitive sequences accounted for 60.41%, and 52.50% of HapA and HapB assemblies, respectively (Table S3). Long terminal repeat (LTR) elements made up the largest

proportion of the repeat sequences. The HapA and HapB exhibited differences in the transposable element (TE) content, particularly LTR sequence composition. *Gypsy* retrotransposons occupied 28.43% of HapA, but only 21.79% of HapB, suggesting asymmetric TE expansion between the two haplotypes. A comprehensive strategy combining transcriptome, *ab initio* and homologue-based predictions were used to annotate the protein-coding genes. Approximately 45,000 genes were annotated in these genomes, with BUSCO values of at least 95.6. (Table 1).

Table 1 Statistics of the HZ genome assemblies and annotations.			
Assembly	Primary	HapA	HapB
Contig size (Mb)	489.22	462.57	449.29
Total no. of contigs	46	69	60
Contig N50 (Mb)	43.17	12.92	16.73
Anchor rate (%)	97.48	94.42	96.09
BUSCO completeness (%)	99.14	97.20	96.60
LTR assembly index (LAI)	17.16	16.23	23.05
Annotation			
Repetitive sequences (%)	61.23	60.41	52.50
Number of annotated genes	45,728	44,775	43,994
Mean mRNA length (bp)	1,930	1,914	1,903
Mean exon length (bp)	248	251	248
Mean exon number per mRNA	5.13	5.11	5.07
BUSCO completeness (%)	98.02	96.21	95.66

3.2 Extensive structural divergence between the two HZ haplotypes

Whole-genome sequence alignment revealed approximately 75% syntenic sequences were syntenic between HapA and HapB (Table S4), whereas numerous structural variations were detected, including inversions, unaligned regions, and local rearrangements (Fig. 2a). Large chromosomal inversions were identified on chromosomes 1 and 5 and confirmed by the Hi-C interaction map (Fig. S3). Additionally, extensive divergence was observed on chromosome 7, where nearly half of the chromosome displayed disrupted collinearity between the two haplotypes. A total of 60.01 Mb and 53.02 Mb of sequences in HapA and HapB, respectively, could not be aligned to the alternative haplome, indicating the presence of

abundant haplotype-specific regions. To validate these divergent regions, we mapped the HiFi reads onto both haplotypes (Fig. S4). Shared genomic regions exhibited an average depth of approximately 52X, whereas haplotype-specific regions showed roughly half the depth (26X), consistent with hemizygous or highly divergent sequences. Despite the extensive sequence divergence, only 1.75% of genes were identified as hemizygous, which is comparable to other self-pollinating species (Peng et al., 2025). While, transposable elements constituted the majority of the haplotype-specific regions, accounting for 81.64% and 67.20% of the unique sequences in HapA and HapB, respectively, which are both higher than their corresponding whole genome levels. These findings suggest that TE expansion and diversification contributed substantially to haplotype differentiation in HZ radish.

3.3 Comparative genomics reveals evolutionary features of HZ radish

To investigate the evolutionary history of the HZ radish, we analyzed orthologous genes across HZ haplotypes, 12 additional radish accessions, as well as the related species *Brassica oleracea*. A total of 5,420 single copy orthologues were identified and used to construct the phylogenetic tree (Fig. 2b). The resulting phylogeny showed that two HZ haplotypes clustered closely together, and were positioned adjacent to the clade containing 'Yanzhong', 'Yuanmou Oli' and 'Xinlimei'.

Gene family identified 728 and 581 haplotype-specific gene families in HapA and HapB, respectively. Functional enrichment analysis revealed that HapA-specific gene families were significantly enriched in biological processes including cell wall biogenesis and hemicellulose metabolic process (Fig. 2c), whereas HapB-specific gene families were enriched in primary metabolic process, gene expression, plant-type cell wall biogenesis, cellulose microfibril organization and so on (Fig. S5). These results suggest that functional divergence between the two haplotypes may partially affect cell wall-related traits in HZ radish.

3.4 Population genomic analyses reveal the unique genetic background of HZ radishes

HZ radishes are traditionally maintained through open pollination. We then further investigated whether the genomic characteristics observed in the genome-assembled accession were common in HZ population. Seven additional HZ radish accessions collected from different cultivation regions were resequenced with an average depth of 94X (Table S1). Genome survey analysis revealed that all these HZ accessions possessed relatively high heterozygosity levels, ranging from 0.68% to 1.04% (Fig. S6), which were substantially higher than those reported for several modern cultivated radish genomes, including 'Xinlimei' (Zhang et al., 2021) and 'NAU-LB' (Xu et al., 2023). Notably, the divergent regions identified on chromosomes 5 and 7 in the assembled HZ genome were also detected in other HZ accessions. Mapping depth analysis revealed obvious coverage differences in these regions among different HZ accessions (Fig. 3a), indicating that these structurally divergent regions are widely maintained within the HZ population. Wei et al., (2022) have identified QTLs associated with root shape

on chromosome 7. We found these QTLs located on the unique regions of HZ haplomes. Thus, those haptype-specific regions on chromosomes 5 and 7 may contribute to the maintenance of the special root shape of HZ radish.

Figure 3 Population analysis of radishes. a. Mapping depth of resequencing reads of seven HZ accessions and HiFi reads of HZ on chromosomes 5 and 7 of HapA and HapB. Window size: 500 kb. The blue and orange boxes show the root-shape QTLs on HapA and HapB, respectively. b. PCA analysis of the *Raphanus* population. c. Phylogenetic tree of the *Raphanus* population. The outer circle indicates the color and shapes of taproots, while the unlabeled ones are other colors and shapes. d. GO enrichment analysis on genes with unique SNPs in HZ radishes.

To further characterize the genetic background of HZ radishes, population genomic analyses were conducted together with 190 published cultivated and wild radish accessions. A total of 5,934,030 high-quality SNPs were identified and used for population structure analysis. PCA analysis showed that all the HZ accessions clustered closely together across the PC1, PC2, or PC3 dimensions (Fig. 3b). Consistently, phylogenetic analysis indicated that all HZ accessions formed an independent clade, and were genetically close to the radishes with elliptic taproots (Fig. 3c). These findings demonstrate that HZ radishes possess a distinct genetic background despite their open-pollinated breeding history.

We then identified HZ-specific SNPs in the HZ radishes, i.e., SNPs present in at least two HZ radishes and absent from all other radishes. 21,004 unique SNPs were identified, and these SNPs mainly located in intergenic region, then the upstream, downstream regions and gene body of 3,787 genes (Table S5). Functional enrichment analysis revealed these genes were significantly enriched in some intracellular metabolic processes like acyl-CoA and thioester metabolic processes, regulation of hydrolase activity, response to environmental pH and aluminum ion (Fig. 3d), suggesting that local environmental adaptation may have contributed to the formation and maintenance of the HZ genetic background.

3.5 Elevated DMC in HZ taproots is associated with enhanced cell wall deposition

HZ radish is widely used for drying and pickling because of its exceptionally high DMC. To investigate the biological basis underlying this trait, DMC was measured in HZ and seven other radish cultivars. Mature HZ taproots exhibited significantly higher DMC than the low DMC cultivars Dachangba (DCB), Qingtong and 791 (Fig. 4a). Since DCB had the lowest DMC, we took it as the representative of low DMC cultivars in the following analyses. Dynamic analysis across four developmental stages showed that HZ consistently maintained higher DMC throughout taproot development (Fig. 4b). Transmission electron microscopy revealed pronounced differences in cell wall ultrastructure between HZ and DCB taproots (Fig. 4c). Quantitative analysis demonstrated that the average cell wall thickness in HZ reached 5.47 μm , which was nearly two-fold higher than that of DCB (2.89 μm) (Fig. 4d).

To determine which cell wall components contributed to elevated DMC in HZ, major structural polysaccharides and fibers were quantified. Cellulose and hemicellulose content was significantly higher

in HZ than in DCB, whereas comparatively smaller differences were observed for lignin (Fig. 4e). In addition, crude fiber content was also elevated in HZ taproots. Together, these results indicate that enhanced cell wall deposition, particularly increased cellulose and hemicellulose accumulation, represent major cytological and biochemical basis underlying the high dry matter phenotype of HZ radish.

3.6 Transcriptomic analyses revealed cellulose biosynthesis was associated with high dry matter accumulation

To investigate the molecular mechanism underlying the higher dry matter accumulation in HZ taproots, comparative transcriptome analyses were performed using the taproots of the DCB and HZ radishes at four developmental stages (Fig. 4b). RNA-seq reads were aligned to the HZ primary assembly to ensure consistent transcript quantification. PCA clustering revealed clear separation between HZ and DCB transcriptomes across developmental stages, indicating substantial transcriptional divergence between the two accessions (Fig. S7). Differential expression analysis revealed 10,776, 14,386, 9,187 and 10,445 differentially expressed genes (DEGs) between the two accessions at the four developmental stages, respectively. Among these, 887 genes were consistently upregulated in HZ across all four stages (Fig. 4f, Table S6). GO enrichment analysis demonstrated that these constitutively upregulated genes were significantly enriched in biological processes such as response to stress, and cellulose biosynthetic and metabolic process (Fig. 4g). Multiple genes involved in secondary cell wall formation and cellulose deposition exhibited markedly higher expression in HZ. These findings suggest that transcriptional activation of cellulose biosynthetic pathways contributes substantially to enhanced cell wall formation and elevated dry matter accumulation in HZ taproots.

3.7 Expression divergence of the cellulose synthase gene superfamily in HZ radish

To further investigate the molecular basis of cellulose accumulation in HZ radish, we systematically identified the cellulose synthase (CESA) and cellulose synthase-like (Csl) gene families in HZ. A total of 15 Cesa and 31 Csl genes were found in HZ, exceeding the corresponding gene numbers in *Arabidopsis thaliana* (Fig. 5a, Table S7). These genes unevenly distributed on the chromosomes (Fig. S8), and five tandemly duplicated Csl gene pairs were identified, indicating that tandem duplication likely contributed to the expansion of Csl genes in HZ.

Expression profiling revealed that most Cesa genes expressed in taproots with FPKM ≥ 1 (Fig. 5b), whereas fewer than half of the Csl genes showed detectable expression (Fig. 5c). Notably, one Cesa (HZch05g017540) and five Csl genes (HZch01g014930, HZch01g014940, HZch01g072460, HZch02g025450 and HZch03g007870) consistently exhibited showed significantly higher expression in HZ than in DCB at all stages. We further validated the expression patterns of three candidate genes at the final developmental stage using RT-qPCR and obtained results consistent with the transcriptomic

data (Fig. 5d). Collectively, these findings suggest that enhanced expression of cellulose synthase-related genes may contribute to the elevated DMC in HZ.

In addition, the cis-regulatory elements within the *CesA* and *CsI* gene promoters (the upstream 2000 bp region) were predicted, and we found numerous cis-regulatory elements associated with plant hormone response elements, including methyl jasmonate, abscisic acid, gibberellin, and salicylic acid responsive elements (Fig. S9). Furthermore, multiple MYB-binding motifs were enriched in *CesA* and *CsI* gene promoters, suggesting potential transcriptional regulation by MYB transcription factors.

To further elucidate the regulatory mechanisms controlling cellulose accumulation in HZ radish, we examined the expression patterns of transcription factors associated with secondary cell wall formation. Among the candidate regulators, three *RsMYB* genes showed significantly higher expression levels in HZ compared to DCB taproots throughout taproot development (Fig. S10). Simultaneously, multiple *CesA* and *CsI* genes involved in cellulose biosynthesis exhibited constitutively elevated expressions in HZ throughout taproot development. Combined with the enrichment of MYB-binding motifs in the promoters of cellulose synthase-related genes, these findings support a regulatory framework in which increased expression of MYB transcript factors may promote the transcriptional activation of downstream cellulose biosynthetic genes. Furthermore, genes associated with secondary cell wall organization, polysaccharide metabolism, and cell wall biogenesis were coordinately upregulated in HZ taproots.

Based on these observations, we propose a regulatory model in which transcriptional reprogramming involving MYB-mediated regulation, activation of cellulose synthase pathways, and enhanced secondary cell wall deposition collectively contributes to elevated cellulose accumulation and high dry matter content in HZ radish taproots (Fig. 5e).

Figure 5 Cellulose synthase gene superfamily and their expression pattern in HZ taproots. a. Cellulose synthase gene superfamily in HZ and Arabidopsis. Expression patterns of *CesA* (b) and *CsI* (c) genes in taproots. d. Relative expression level of three cellulose biosynthetic genes by RT-qPCR. Data are presented as mean $\pm$ SD. Statistical significance was determined using Student's *t*-test (*, $p < 0.05$; **, $p < 0.01$). e. Proposed regulatory model underlying elevated cellulose accumulation and high dry matter formation in HZ radish

4. Discussion

In this study, we present the haplotype-resolved genome assemblies for *R. sativus* 'Huangzhou', a landrace radish in China, combining PacBio HiFi, ONT ultra long and Hi-C data. Compared with previously published radish genomes, the HZ genome exhibited high assembly continuity and completeness, providing a valuable genomic resource for future studies on radish evolution, domestication, and trait improvement.

One notable feature of the HZ genome is its relatively high heterozygosity. The estimated heterozygosity rate of HZ (1.34%) was substantially higher than those reported cultivated radish genomes, including

'Xinlimei', 'NAU-LB', and 'C60213' (Zhang et al., 2021; Xu et al., 2023; Yang et al., 2025). In contrast, the heterozygosity level of HZ was more comparable to that observed in certain wild or weedy radishes (Zhang et al., 2021). This pattern likely reflects the long-term maintenance of HZ radishes through traditional open-pollinated cultivation systems. Haplotype-resolved assembly further revealed extensive structural divergence HapA and HapB, especially on chromosomes 5 and 7. Most haplotype-specific regions were enriched with transposable elements, suggesting that TE expansion and diversification played important roles in shaping haplotype differentiation in HZ radish. Similar TE-associated structural divergence has been widely reported in highly heterozygous crop genomes and is increasingly recognized as an important driving force for phenotypic diversification and environmental adaptation (Tao et al., 2025). Interestingly, several divergent regions identified in HZ overlapped with previously reported root-shape QTLs (Wei et al., 2022), suggesting that haplotype-specific structural variation may contribute to the maintenance of HZ characteristics, like root shape. Furthermore, resequencing analyses showed that the divergence of haplotypes were commonly present in geographically separated HZ accessions, indicating that they may represent stable genomic features associated with landrace-specific traits.

Landraces are important reservoirs of genetic diversity because they have undergone long-term adaptation to local environments. This adaptability is essential for their survival and contribution to agricultural sustainability. Population genomic analyses demonstrated that HZ radishes formed an independent genetic cluster despite their open-pollinated breeding history, supporting the existence of a relatively conserved genetic background within the HZ population. We identified more than 21,000 HZ-specific SNPs associated with genes enriched in environmental responses, including responses to environmental pH and aluminum ion stress. Some cultivation regions in Hubei Province have encountered the issue of soil acidification, where aluminum toxicity may influence crop growth and productivity. Therefore, these enriched pathways may reflect local environmental adaptation during long-term cultivation of HZ radishes. The relatively high genomic diversity observed in HZ radishes may provide important advantages for breeding and environmental resilience. Modern crop improvement has often reduced genetic diversity through repeated artificial selection and inbreeding, whereas many landraces possess adaptive alleles that are absent from cultivars (Khoury et al., 2022). Our findings highlight the importance of conserving and utilizing traditional landrace germplasm for future radish breeding.

Dry matter accumulation is a key determinant of taproot quality, texture, and processing suitability in radish. HZ radish is widely used for drying and pickling because of its high DMC and characteristic texture. However, the biological basis underlying this trait has remained largely unclear. The study of dry matter including cellulose in HZ radish provides clues for breeding of new varieties that are better suited for both fresh eating processing. Combined with cytological, biochemical, and transcriptomic analyses, we demonstrated that enhanced cell wall deposition, particularly cellulose accumulation, is strongly associated with the high DMC of HZ taproots. Ultrastructural observations revealed thicker cell walls in HZ compared with the low-dry-matter cultivar DCB, and biochemical analyses showed that cellulose significantly increased. Previous studies have reported that cell wall composition strongly influences

texture and processing quality in radish and other root vegetables. Our findings further suggest that cellulose accumulation is a major contributor to dry matter formation in HZ radish, which was similar to the result on metabolomic profiles for assessment of the nutritional value of processed radish (Wang et al., 2023). Increased cellulose deposition may enhance mechanical strength and tissue compactness, thereby contributing to the desirable processing characteristics of HZ taproots.

Cellulose biosynthesis is controlled by a complex transcriptional regulatory network involving cellulose synthase complexes and secondary wall-associated transcription factors, like NAC and MYB (Wang et al., 2012; Pedersen et al., 2023). A NAC transcription factor, *RsSND1*, was found to regulate biosynthesis of secondary cell wall in radish (Wang et al., 2024). In this study, we identified *CesA* and *CsI* gene families in the HZ genome and found that multiple cellulose synthase-related genes were constitutively upregulated in HZ taproots. Among the candidate regulators, three *RsMYB* genes showed significantly increased expression in HZ compared with DCB during early taproot development. MYB transcription factors are well-established regulators of secondary cell wall formation (Geng et al., 2020; Qiang et al., 2022; Wang et al., 2019; Xiao et al., 2021). Arabidopsis MYB46 and MYB83 were found to be the second layer of the main switch of secondary cell-wall biosynthesis, which coordinated upstream and downstream secondary wall synthesis-related transcription factors, including MYB genes. The enrichment of MYB-binding motifs in the promoters of *CesA* and *CsI* genes further supports the existence of MYB-mediated transcriptional regulation in HZ radish. Therefore, we propose that increased expression of *RsMYBs* may release transcriptional repression on downstream cellulose biosynthetic genes, thereby promoting cellulose accumulation and secondary cell wall deposition in HZ taproots. Although additional functional validation is still required, our integrative analyses provide a preliminary regulatory framework linking transcriptional regulation, cellulose biosynthesis, secondary cell wall formation, and dry matter accumulation in radish taproots.

5. Conclusions

In this study, we generated a high-quality haplotype-resolved genome assembly of the landrace radish ‘Huangzhou’ and systematically investigated its genomic diversity and population characteristics. Integration of cytological, biochemical, and transcriptomic analyses demonstrated that enhanced cellulose accumulation and secondary cell wall deposition contributed to the elevated dry matter content of HZ taproots. Multiple cellulose synthase-related genes were consistently upregulated in HZ, and increased expression of the *RsMYB* transcription factors may participate in transcriptional regulation of cellulose biosynthesis pathways. Overall, our study provides valuable genomic resources for radish research and offers new insights into the genetic and molecular basis of cellulose accumulation and dry matter formation in landrace radishes.

Abbreviations

CSC cellulose synthase complex

CESA cellulose synthase

DEGs differentially expressed genes

DMC dry matter content

EDTA Extensive de novo TE Annotator

GSA Genome Sequence Archive

Hi-C high-throughput chromatin conformation capture

KEGG Kyoto Encyclopedia of Genes and Genomes

LAI LTR Assembly Index

LTR long terminal repeat

ONT Oxford Nanopore Technologies

PacBio Pacific Biosciences

PCW primary cell wall

RPII RNA polymerase II

SCW secondary cell wall

SNP single nucleotide polymorphism

TE transposable element

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The genome sequencing and assembly data have been deposited in the Genome Sequence Archive (GSA) database at BIG Data Center (<https://ngdc.cncb.ac.cn/gsa/>) under BioProject Number

PRJCA059937. Additionally, the assembled genome and annotations have been submitted to FigShare for public access (<https://figshare.com/account/articles/31703188>).

Competing Interests

All authors declare no conflict of interest.

Funding Declaration

This work was supported by the Joint Project of Hubei Provincial Natural Science Foundation (2026AFC0441) .

Authors' contributions

S.L. and K.X. conceived and designed the research. X.Y. and H.L. carried out the bioinformatics analysis. S.L., K.X., Y.L., Y.T., W.T., H.D., C.G. and X.Y. conducted experiments and analyzed data. S.L. and X.Y. wrote the manuscript. S.L., H.X., L.G. and XY. revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the Hubei Special Project for Development of Science and Technology in Local by Central Guidance (2024CSA071).

References

1. Bisma A, Lucas CR, Riedl KM, Clinton K, Mortazavi A. Cruciferous vegetables, isothiocyanates and bladder cancer prevention. *Mol Nutr Food Res*. 2018;62:1800079.
2. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18(4):366–8.
3. Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinform*, 48(1):4.11.1–4.11.39.
4. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90.
5. Cheng H, Asri M, Lucas J, Koren S, Li H. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat methods*. 2024;21(6):967–70.
6. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;1000(15):2156–8.
7. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB,

- Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491–8.
8. Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protoc.* 2007;2(9):2233–44.
9. Do MH, Lee HB, Oh MJ, Jhun HJ, Choi SY, Park H. Polysaccharide fraction from greens of *Raphanus sativus* alleviates high-fat diet-induced obesity. *Food Chem.* 2021;343:128395.
10. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* 2017;356(6333):92–5.
11. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 2016;3(1):95–8.
12. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238.
13. Fan L, Wang Y, Xu L, Tang M, Zhang X, Ying J, Li C, Dong J, Liu L. A genome-wide association study uncovers a critical role of the *RsPAP2* gene in red-skinned *Raphanus sativus* L. *Hortic Res.* 2020;7(1):164.
14. Fu J, Jiang J, Ge C, Cao W, Xiong X, Li S. Comparative analysis on the different developmental stages taproots of ‘Huangzhou’ radish. *Mol Plant Breed.* 2022;20(14):4772–8.
15. Geng P, Zhang S, Liu J, Zhao C, Wu J, Cao Y, Fu C, Han X, He H, Zhao Q. MYB20, MYB42, MYB43, and MYB85 regulate phenylalanine and lignin biosynthesis during secondary cell wall formation. *Plant Physiol.* 2020;182(3):1272–83.
16. Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 2019;20(1):277.
17. Hanlon PR, Webber DM, Barnes DM. Aqueous extract from Spanish black radish (*Raphanus sativus* L. var. *niger*) induces detoxification enzymes in the HepG2 human hepatoma cell line. *J Agric Food Chem.* 2007;55(16):6439–46.
18. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods Mol Biol*, 1962:65–95.
19. Hu H, Wang J, Hu T, Wang Y, Li Y, Wang W, Wei Q, Yan Y, Bao C. 2022. Research progress on radish genetic breeding during The Thirteenth Five-year Plan in China. *China Vegetables*, (10):20–6.
20. Hur GH, Lee TK, Cho Y-J, Kim JH, Park JHY, Yang H, Lee KW. Optimization of cultivar, germination time, and extraction for radish sprout extract with high sulforaphene content. *J Sci Food Agric.* 2024;104:5010–20.
21. Jiang Q, Zhao S, Zhao W, Wang P, Qin P, Wang J, Zhao Y, Ge Z, Zhao X, Wang D. The role of water distribution, cell wall polysaccharides, and microstructure on radish (*Raphanus sativus* L.) textural properties during dry-salting process. *Food Chem X.* 2024;22:101407.

22. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–15.
23. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 2019;20(1):278.
24. Lee SO, Lee IS. Induction of quinone reductase, the phase 2 anticarcinogenic marker enzyme, in Hepa1c1c7 cells by radish sprouts (*Raphanus sativus* L). *J Food Sci.* 2006;71:S144–8.
25. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 2024;52:W78–82.
26. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. 1000 Genome Project Data Processing Subgroup.
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
30. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021;38(10):4647–4654.
31. Manivannan A, Kim JH, Kim DS, Lee ES, Lee H-E. Deciphering the nutraceutical potential of *Raphanus sativus*: a comprehensive overview. *Nutrients.* 2019;11(2):402.
32. Miyamoto T, Tobimatsu Y, Umezawa T. MYB-mediated regulation of lignin biosynthesis in grasses. *Curr Plant Biol.* 2020;24:100174.
33. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 2018;46(21):e126.
34. Peng Y, Wang Y, Liu Y, Fang X, Cheng L, Long Q, Su D, Zhang T, Shi X, Xu X, Xu Q, Wang N, Zhang F, Liu Z, Xiao H, Yao J, Tian L, Hu W, Chen S, Wang H, Zhou Y. The genomic and epigenomic landscapes of hemizygous genes across crops with contrasting reproductive systems. *Proc Natl Acad Sci U S A.* 2025;122(6):e2422487122.
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
36. Qiang Z, Sun H, Ge F, Li W, Li C, Wang S, Zhang B, Zhu L, Zhang S, Wang X, Lai J, Qin F, Zhou Y, Fu Y. The transcription factor ZmMYB69 represses lignin biosynthesis by activating ZmMYB31/42 expression in maize. *Plant Physiol.* 2022;189(4):1916–9.

37. Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* 2018;6(2):256–e2581.
38. Slavin J. Fiber and prebiotics: mechanisms and health benefits. *Nutrients.* 2013;5:1417–35.
39. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
40. Su W, Ou S, Hufford MB, Peterson T. A tutorial of EDTA: Extensive de novo TE Annotator. *Methods Mol Biol.* 2021;2250:55–67.
41. Taylor NG. Cellulose biosynthesis and deposition in higher plants. *New Phytol.* 2008;178(2):239–52.
42. Wang HZ, Dixon RA. On-off switches for secondary cell wall biosynthesis. *Mol Plant.* 2012;5(2):297–303.
43. Wang J, Wei Q, Wang W, Hu H, Yan Y, Wang Y, Li Y, Jiang Y, Wu G, Hu T, Bao C. Understanding the nutraceutical diversity through comparative analysis of taproot metabolomes of different edible radish types via UHPLC-Q-TOF-MS. *Food Chem.* 2023;403:134010.
44. Wang L, Lu W, Ran L, Dou L, Yao S, Hu J, Fan D, Li C, Luo K. R2R3-MYB transcription factor MYB6 promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in *Populus tomentosa*. *Plant J.* 2019;99:733–51.
45. Wang L, Hart BE, Khan GA, Cruz ER, Persson S, Wallace IS. Associations between phytohormones and cellulose biosynthesis in land plants. *Ann Bot.* 2020;126(5):807–24.
46. Wang Y, Wang Q, Sun H, Zhang Z, Qian H, Zhao X, He H, Zhang L. Glucosinolate profiles in different organs of 111 radish accessions and candidate genes involved in converting glucobrassicin to 4-hydroxyglucobrassicin. *J Agric Food Chem.* 2022;70:488–97.
47. Wang Y, Wang Q, Wu X, Pang Y, Guo Y, Li Z, Zhang L. A NAC transcription factor RsSND1 regulating secondary cell wall deposition is involved in fleshy taproot formation in radish (*Raphanus sativus* L.). *J Plant Growth Regul.* 2024;43(6):1844–57.
48. Wilson TH, Kumar M, Turner SR. The molecular basis of plant cellulose synthase complex organisation and assembly. *Biochem Soc Trans.* 2021;49(1):379–91.
49. Xiao R, Zhang C, Guo X, Li H, Lu H. MYB transcription factors and its regulation in secondary cell wall formation and lignin biosynthesis during xylem development. *Int J Mol Sci.* 2021;22:3560.
50. Xu L, Wang Y, Dong J, Zhang W, Tang M, Zhang W, Wang K, Chen Y, Zhang X, He Q, Zhang X, Wang K, Wang L, Ma Y, Xia K, Liu L. A chromosome-level genome assembly of radish (*Raphanus sativus* L.) reveals insights into genome adaptation and differential bolting regulation. *Plant Biotechnol J.* 2023;21(5):990–1004.
51. Xu Y, Zhu X, Gong Y, Xu L, Wang Y, Liu L. Evaluation of reference genes for gene expression studies in radish (*Raphanus sativus* L.) using quantitative real-time PCR. *Biochen Bioph Res Co.* 2012;424:398–403.
52. Yang F, Peng S, Yuan S, Ran M, Li X, Li Y, Liu B, Li M, Kong C, Yang X, Pan G, Yong X, Ran K, Kuang N, Zhang D, Lin H. 2025. A telomere-to-telomere genome assembly of radish (*Raphanus sativus* L.)

- provides insights into QTL mapping of bolting traits. *J Genet Genomics*, 7:S1673-8527(25)00222-X.
53. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
54. Zhang S, Ma SH, Wei Y, Liu Y, Dou D, Cui Y, Liang H, Liesche B, Li J, Chen J, S. Mutation of CESA1 phosphorylation site influences pectin synthesis and methylesterification with a role in seed development. *J Plant Physiol*. 2022;270:153631.
55. Zhang X, Liu T, Wang J, Wang P, Qiu Y, Zhao W, Pang S, Li X, Wang H, Song J, Zhang W, Yang W, Sun Y, Li X. Pan-genome of *Raphanus* highlights genetic variation and introgression among domesticated, wild, and weedy radishes. *Mol Plant*. 2021;14(12):2032–55.

Figures

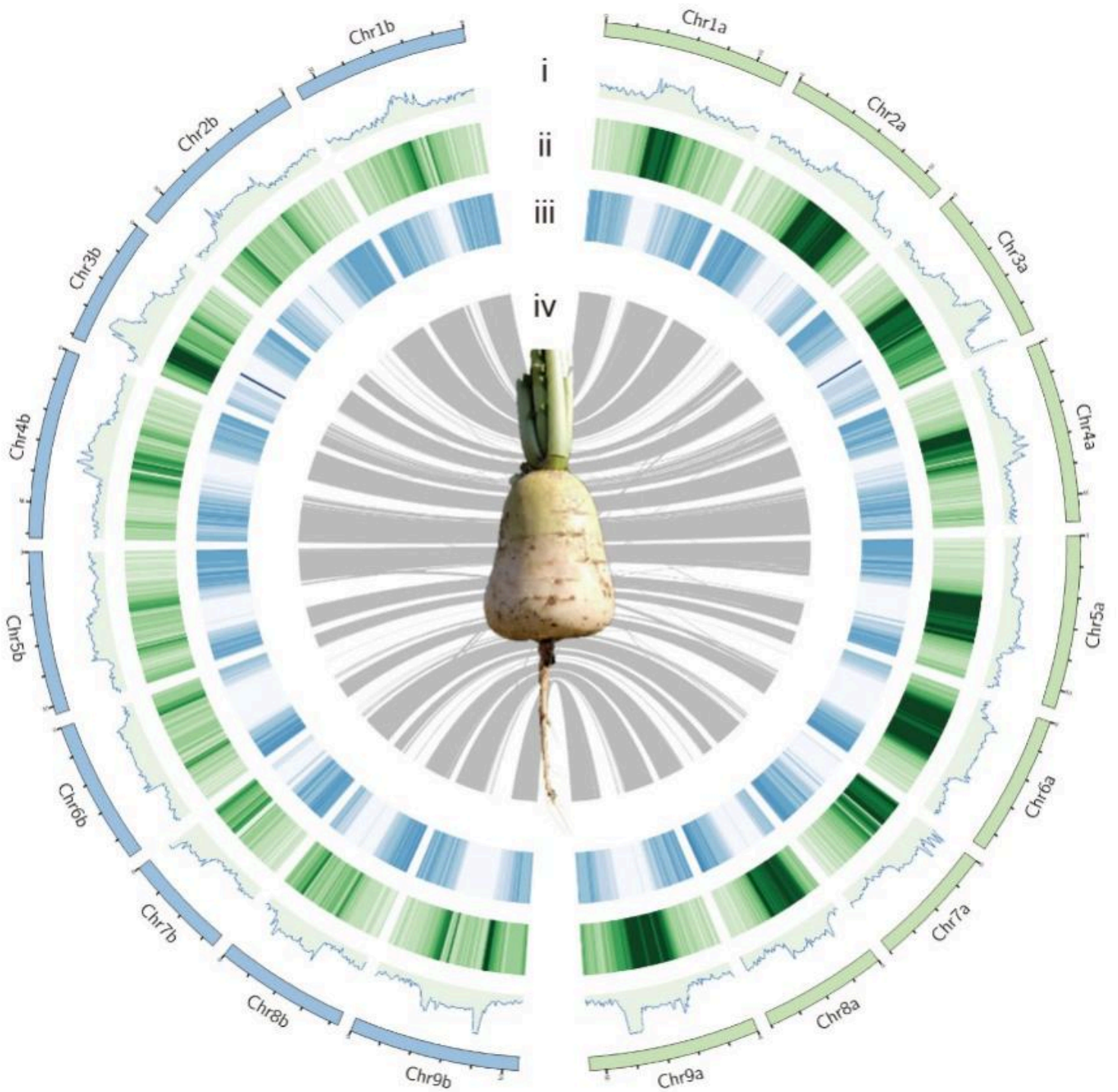


Figure 1

Morphology and genomic features of HZ radish. GC content (i), repeat content (ii), gene density (iii) in 500-kb sliding windows, and chromosomal synteny between haplotypes (iv) are shown. Root type of HZ radish is shown.

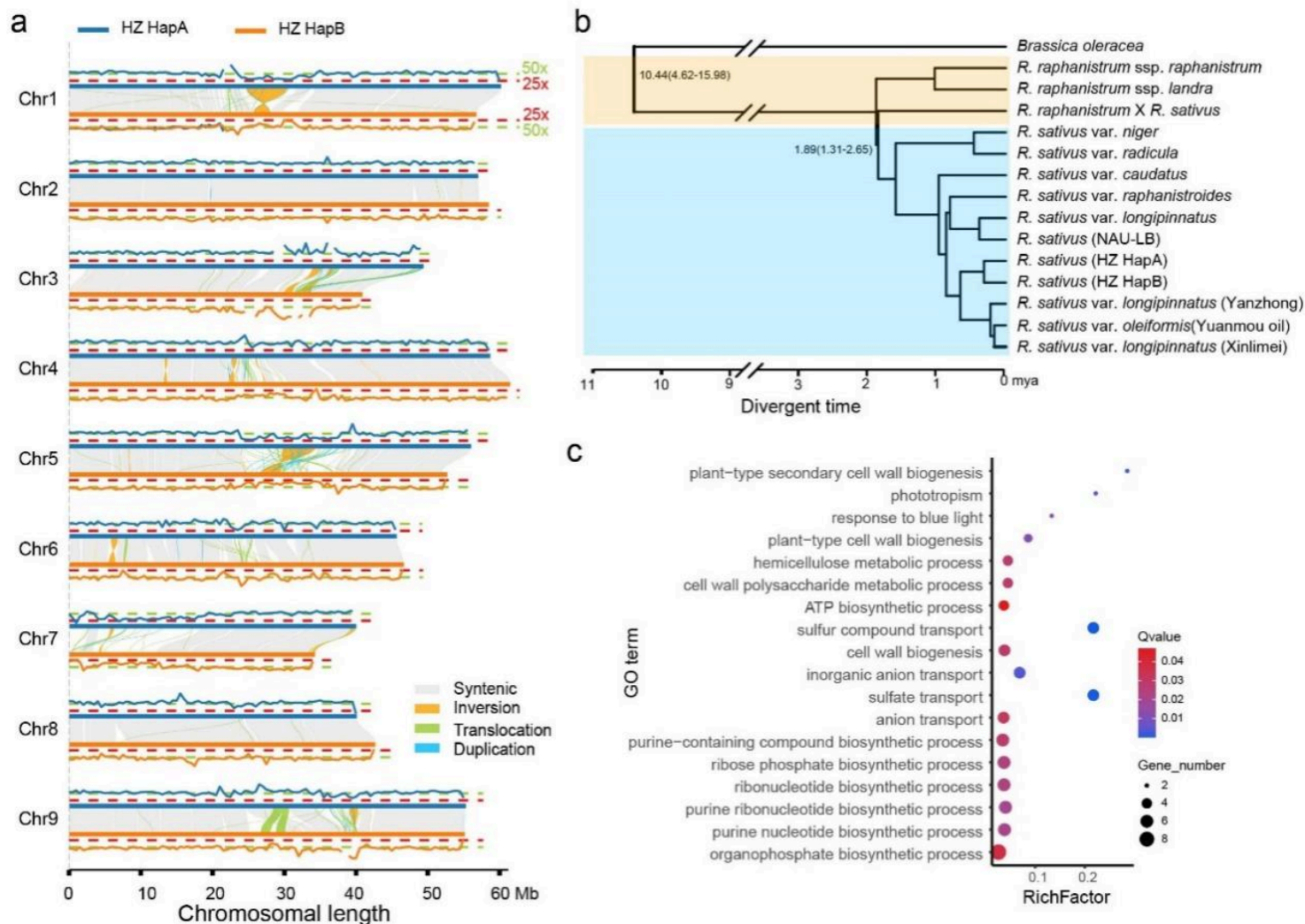


Figure 2

Comparative genomic analyses reveal extensive haplotype divergence in HZ radish. a. Genomic syntenic analysis of the HapA and HapB of genomes and the mapping depth of HiFi reads on them, respectively. Window size: 500 kb. b. Phylogenetic tree based on single copy gene families in HZ and 13 other additional plant species. c. GO enrichment analysis of gene families unique in HapA.

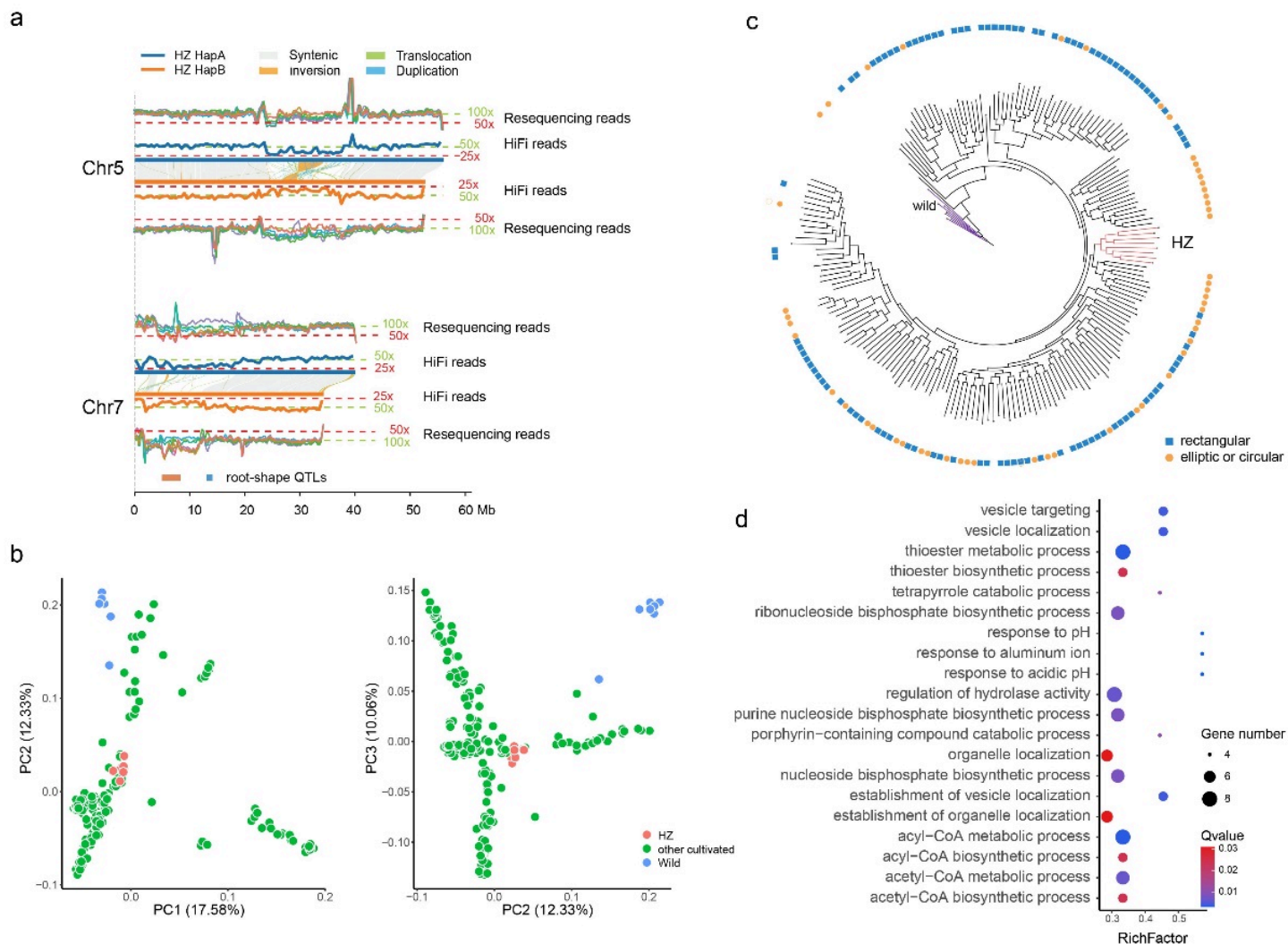


Figure 3

Population analysis of radishes. a. Mapping depth of resequencing reads of seven HZ accessions and HiFi reads of HZ on chromosomes 5 and 7 of HapA and HapB. Window size: 500 kb. The blue and orange boxes show the root-shape QTLs on HapA and HapB, respectively. b. PCA analysis of the *Raphanus* population. c. Phylogenetic tree of the *Raphanus* population. The outer circle indicates the color and shapes of taproots, while the unlabeled ones are other colors and shapes. d. GO enrichment analysis on genes with unique SNPs in HZ radishes.

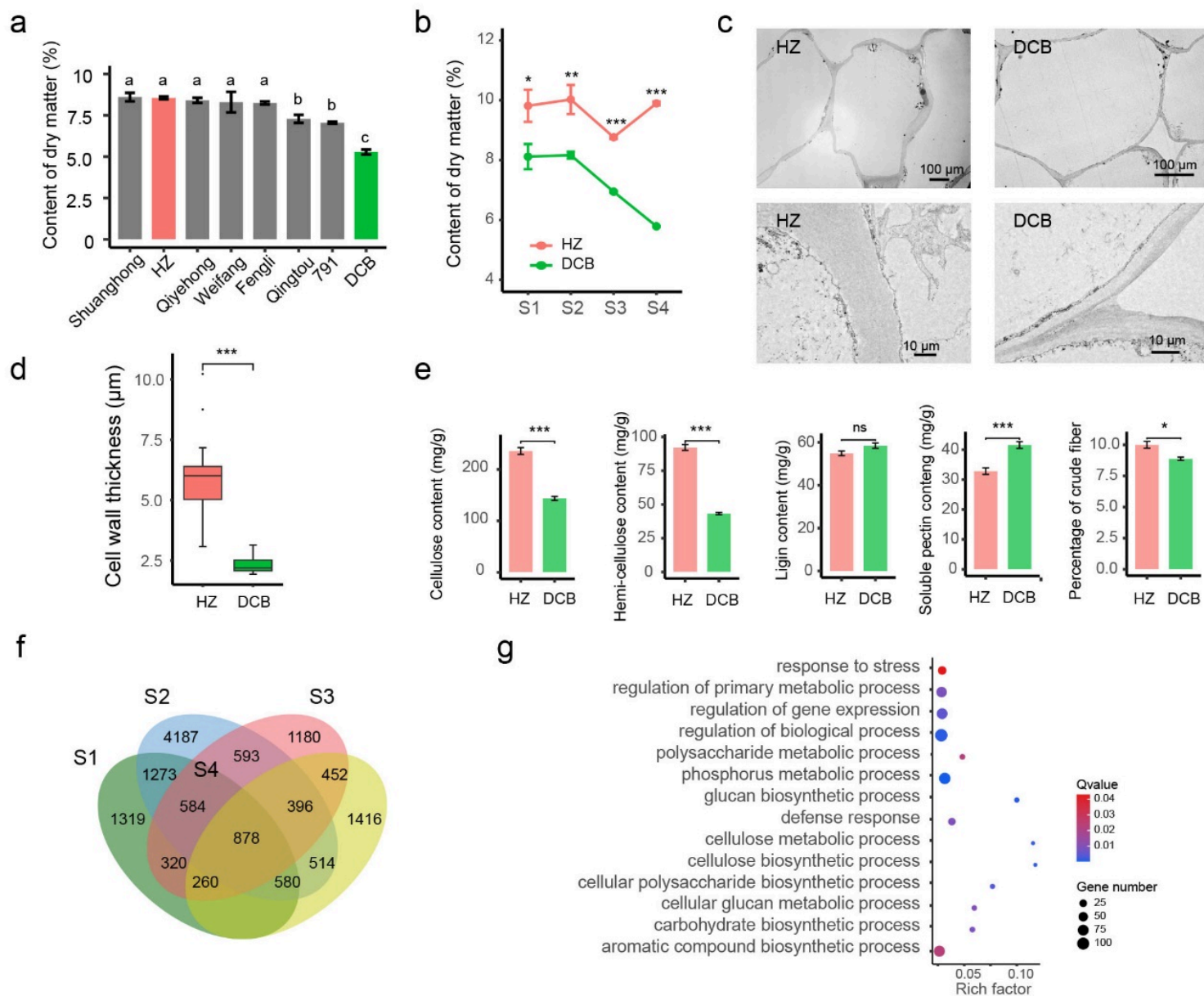


Figure 4

Enhanced cellulose deposition in HZ taproots and transcriptomic analyses. **a**. Dry matter content (DMC) in representative radish cultivars. Statistical significance was determined using one-way ANOVA followed by Duncan's multiple range test. **b**. DMC at four developmental stages of taproots. **c**. Cell wall ultrastructural view of cells in taproots from HZ and DCB. **d**. Quantification of cell wall thickness of HZ and DCB taproots. **e**. Comparative analysis on content of cell wall-associated components, including cellulose, hemicellulose, lignin, soluble pectin and crude fiber. Data are presented as mean $\pm$ SD. Statistical significance was determined using Student's *t*-test (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). **f**. Venn diagram showing differentially expressed genes consistently upregulated in HZ during taproot development. **g**. GO enrichment analysis of constitutively upregulated genes in HZ.

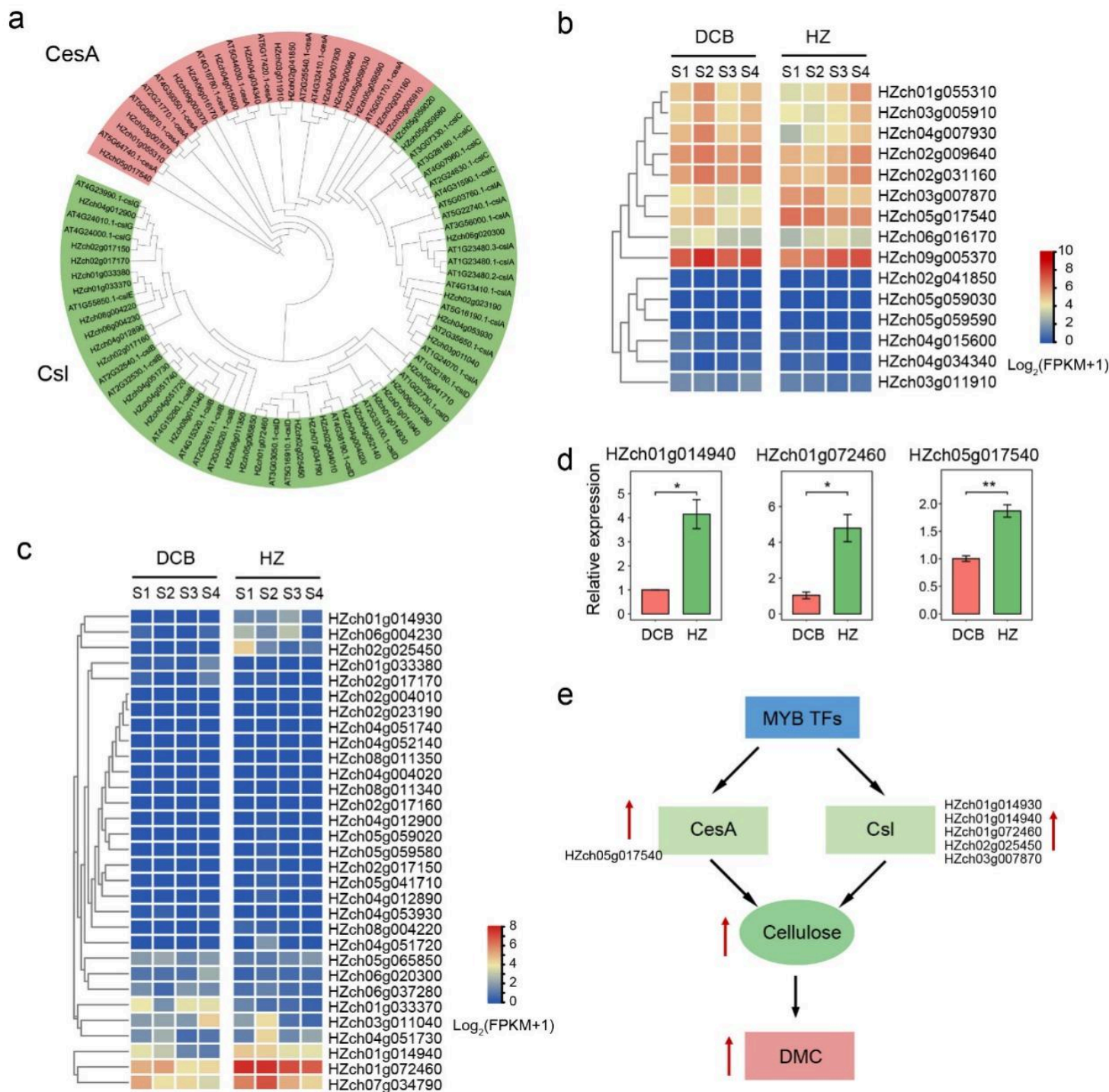


Figure 5

Cellulose synthase gene superfamily and their expression pattern in HZ taproots. a. Cellulose synthase gene superfamily in HZ and Arabidopsis. Expression patterns of *CesA* (b) and *Csl* (c) genes in taproots. d. Relative expression level of three cellulose biosynthetic genes by RT-qPCR. Data are presented as mean $\pm$ SD. Statistical significance was determined using Student's *t*-test (*, $p < 0.05$; **, $p < 0.01$). e. Proposed regulatory model underlying elevated cellulose accumulation and high dry matter formation in HZ radish

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfigures.docx](#)
- [Supplementarytable.xlsx](#)