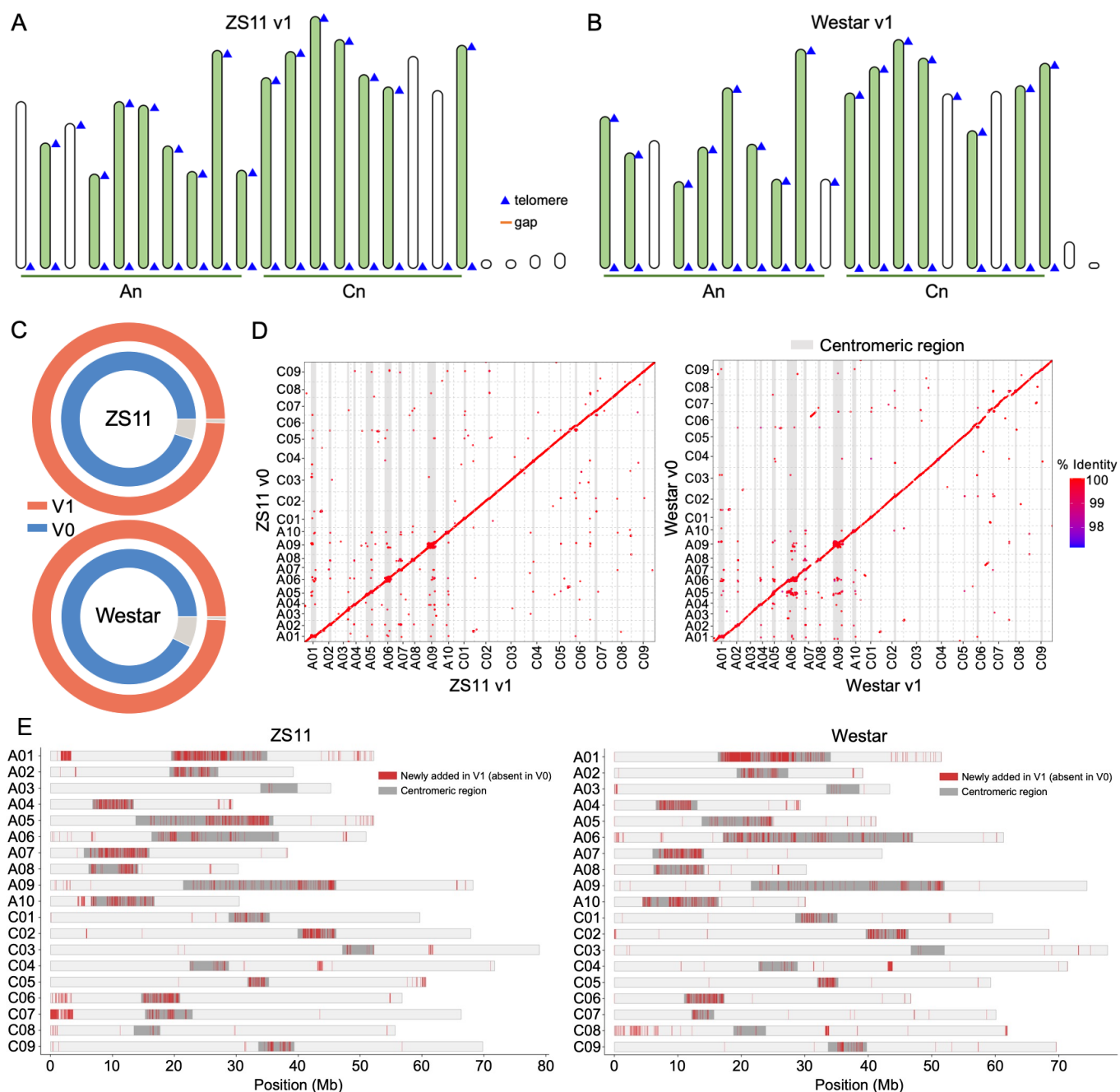
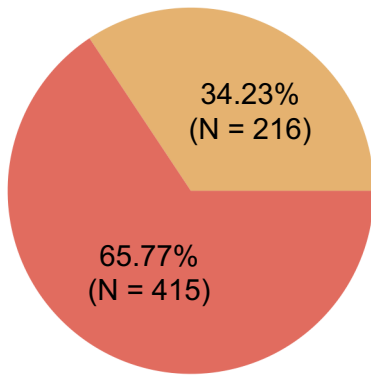
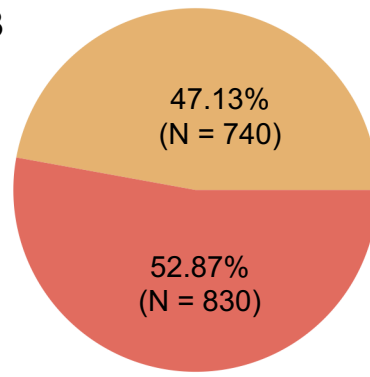
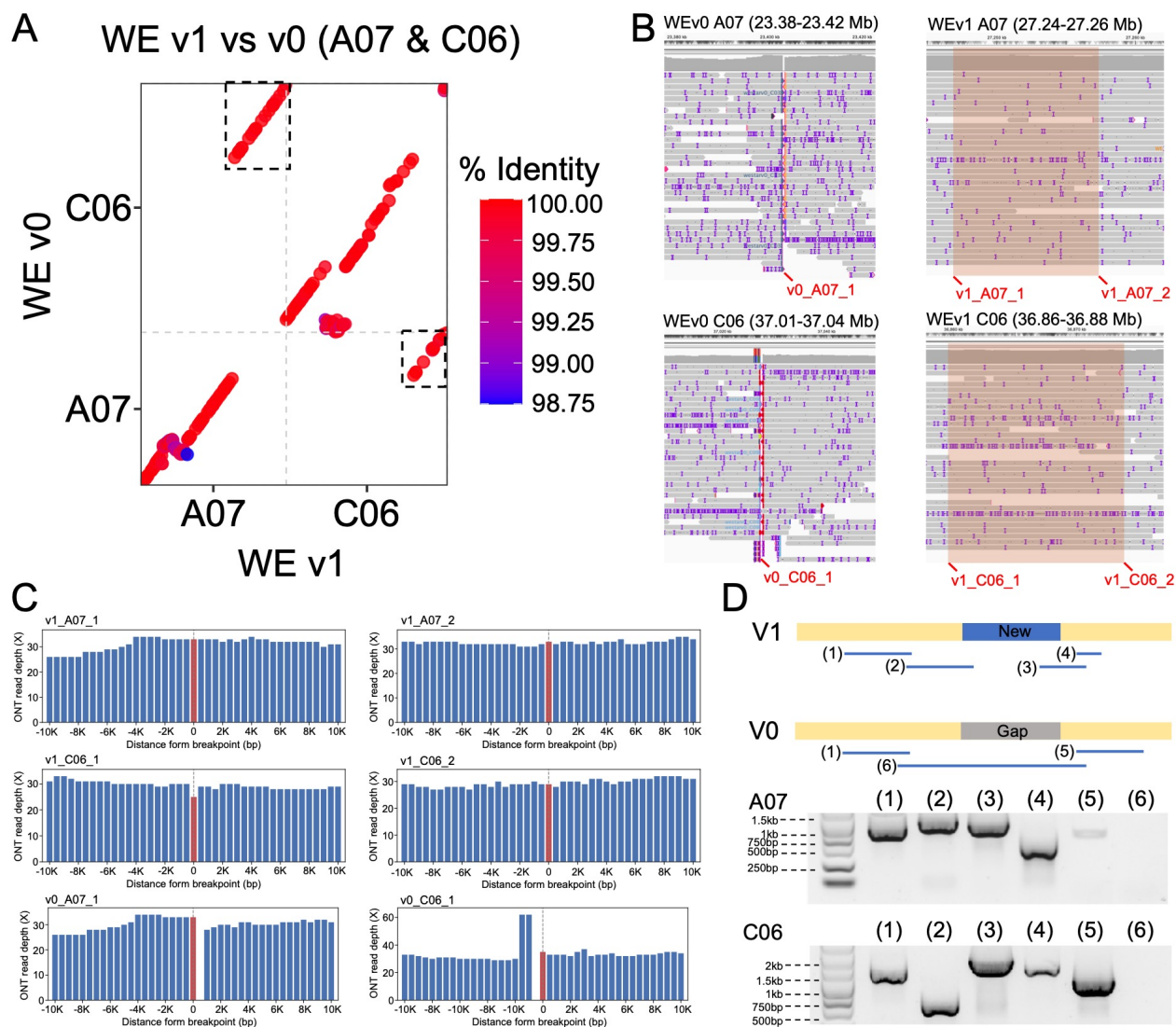




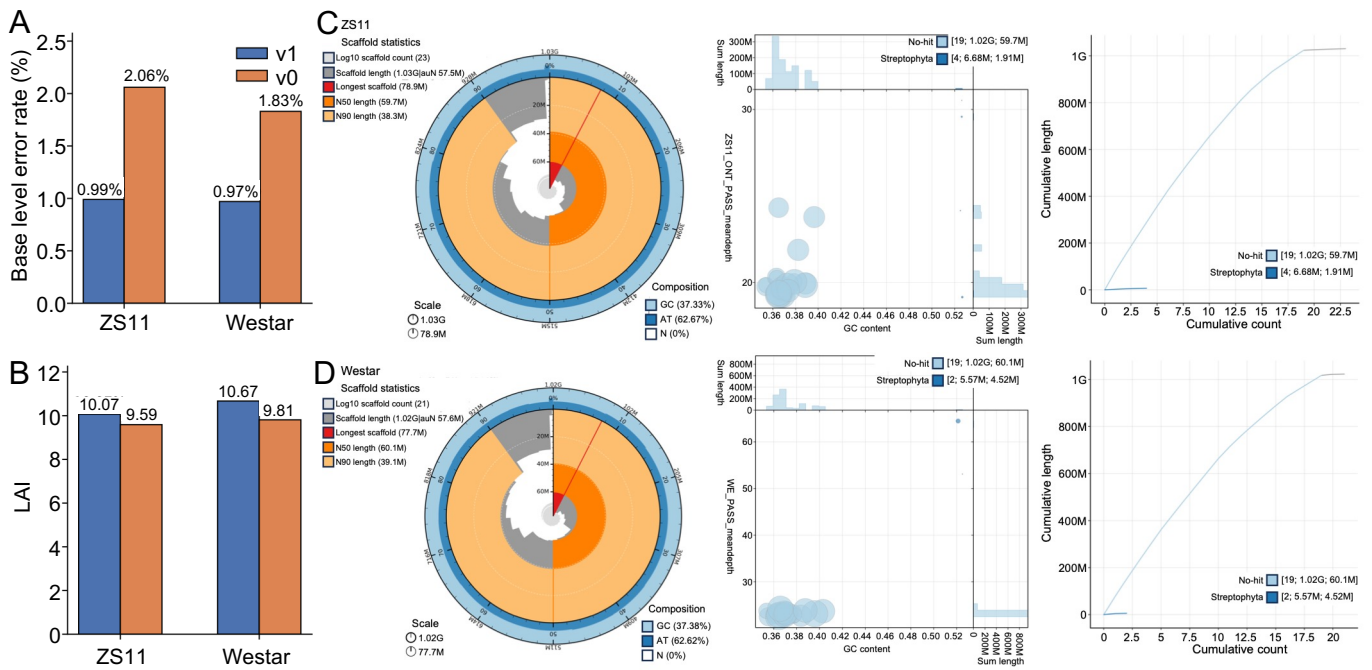
Supplementary Fig 1. Comparison of different versions for ZS11 and Westar. **A-B.** Distribution of telomeres and gaps in the ZS11 v1 and Westar v1 genomes. (A) is the ZS11 v1 genomes. (B) is the Westar v1 genomes. Green highlights the chromosomes that have achieved Telomere-to-Telomere (T2T) continuity. **C.** Chromosomal anchoring rates across different assembly versions. Grey indicates the proportion of unanchored sequences, while orange and blue represent the proportions of sequences anchored to chromosomes in the v1 and v0 genomes, respectively. **D.** Synteny analysis between the different assembly versions of ZS11 and Westar. Grey shaded areas indicate the centromeric regions in the v1 genomes. **E.** Chromosomal distribution of v1-specific sequences identified by whole-genome comparison with the corresponding v0 assemblies in ZS11 and Westar. Red regions indicate sequences newly assembled in v1, whereas gray regions represent centromeric regions.

A**B**

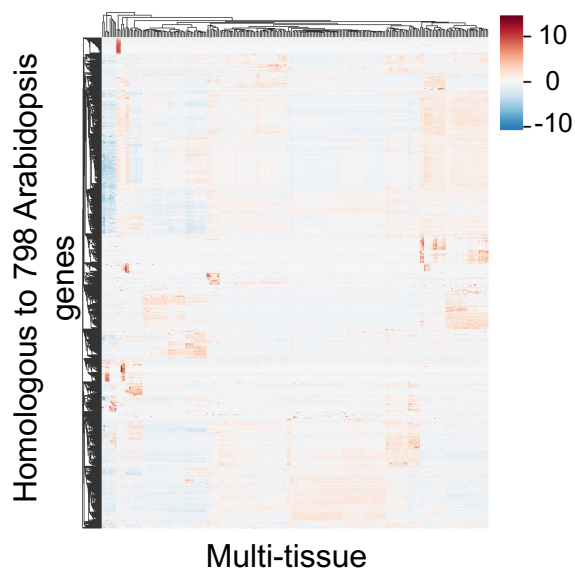
Supplementary Fig 2. Number and proportion of genes expressed in the centromere region of the ZS11 genome. (A) is ZS11 v0 genome, and **(B)** is ZS11 v1 genome. Yellow denotes the proportion of genes with expression (TPM > 1 in at least three samples), while red denotes the proportion of genes without expression.



Supplementary Fig 3. Multi-line evidence validating the structural discrepancy between Westar v1 and v0 assemblies across the A07-C06 region. **A.** Genome-wide synteny dotplot comparing Westar v1 and v0, focused on chromosomes A07 and C06. Point color indicates sequence identity of the aligned segments. Aside from the main diagonal, two off-diagonal alignment blocks (dashed boxes) indicate an interchromosomal structural discrepancy involving chromosomes A07 and C06. **B.** IGV visualization of raw ONT ultra-long reads aligned across the six candidate breakpoint positions. The left panels show the two v0-specific mis-joined junctions, with red dashed lines indicating the breakpoint positions on A07 and C06. Left column: corresponding loci in the v0 assembly, with red dashed lines marking the breakpoint positions (v0_A07_1, v0_C06_1). Right column: corresponding loci in the v1 assembly, with orange-shaded regions indicating sequences newly recovered in v1, flanked by red dashed lines marking the two junction breakpoints. **C.** Bar plots of ONT sequencing depth within ± 10 kb of each of the six candidate junction breakpoints (500-bp bins on the x-axis, showing distance from the breakpoint in bp; y-axis, ONT read depth in $\times$). Red bars indicate the bin containing the breakpoint (distance = 0). **D.** Schematic representation of the breakpoint-specific PCR strategy and the corresponding agarose-gel electrophoresis results. The upper schematics depict the v1 genome structure (newly added sequence segment labeled “New” in blue) and the v0 genome structure (gap segment labeled “Gap” in gray). The lower panels show PCR products generated using primer pairs designed to distinguish the v1 and v0 configurations in the A07 and C06 regions. Amplicons of the expected sizes were obtained for the v1-supported junctions, whereas the corresponding v0-specific junctions were not supported. Source data are provided as a Source Data file.



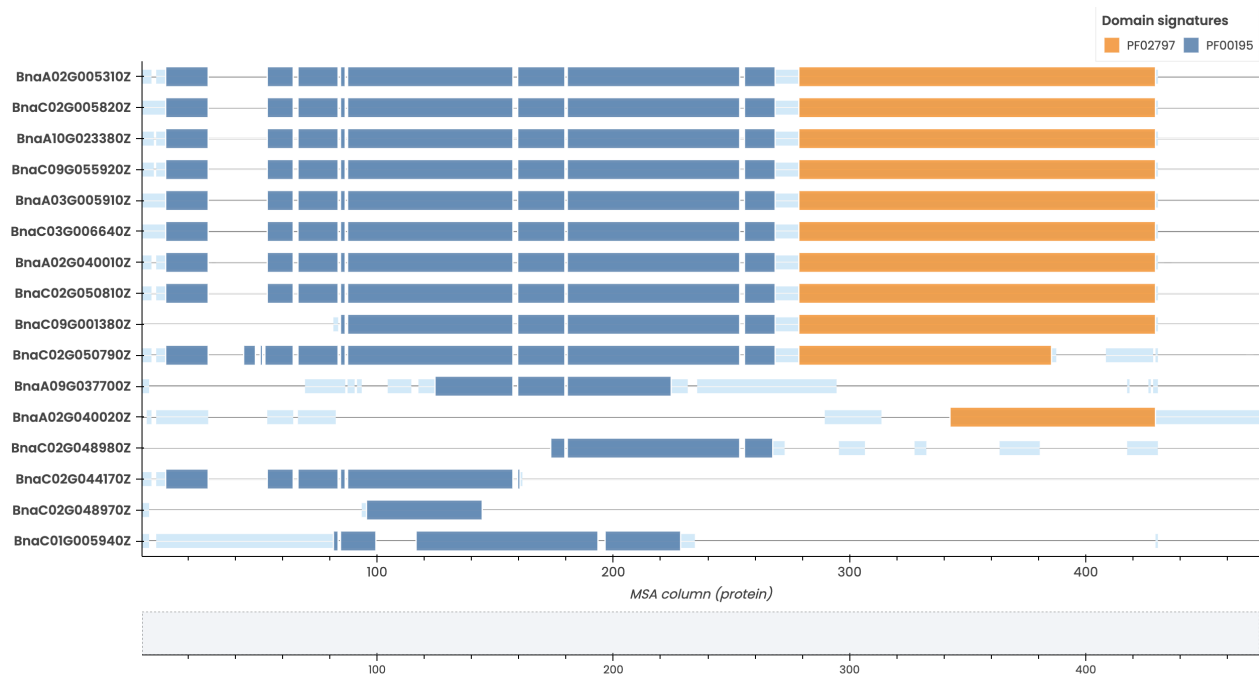
Supplementary Fig 4. Multi-dimensional quality assessment of the v1 and v0 genome assemblies of ZS11 and Westar. **A.** Base-level error rate (%) of the ZS11 and Westar v1 and v0 assemblies, estimated by remapping raw ONT long reads to each assembly. **B.** LTR Assembly Index (LAI) of the ZS11 and Westar v1 and v0 assemblies, reflecting the assembly continuity of repetitive and intergenic regions. **C-D.** BlobTools contamination screening for the ZS11 (C) and Westar (D) v1 assemblies. Left: snail plot summarizing assembly statistics. Middle: GC content–coverage blob plot, where each circle represents a scaffold (circle size proportional to scaffold length) colored by taxonomic assignment based on BLAST hits against the NCBI nucleotide database; side histograms show the length distribution of scaffolds along the GC content and sequencing coverage axes. Right: cumulative assembly length plotted against the number of scaffolds ranked by taxonomic category.



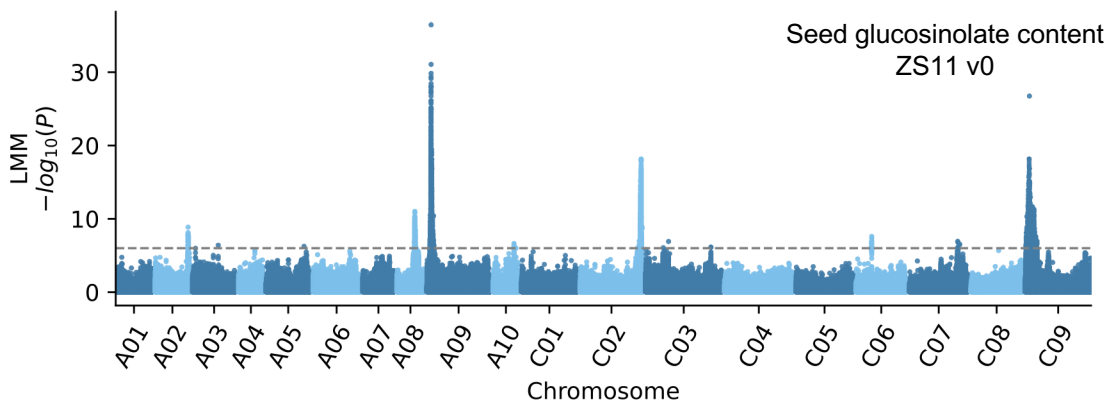
Supplementary Fig 5. Expression heatmap of genes homologous to 798 *Arabidopsis thaliana* genes across multiple tissues. Heatmap colors indicate Z-score-scaled gene expression levels, with blue representing lower expression and red representing higher expression. Source data are provided as a Source Data file.



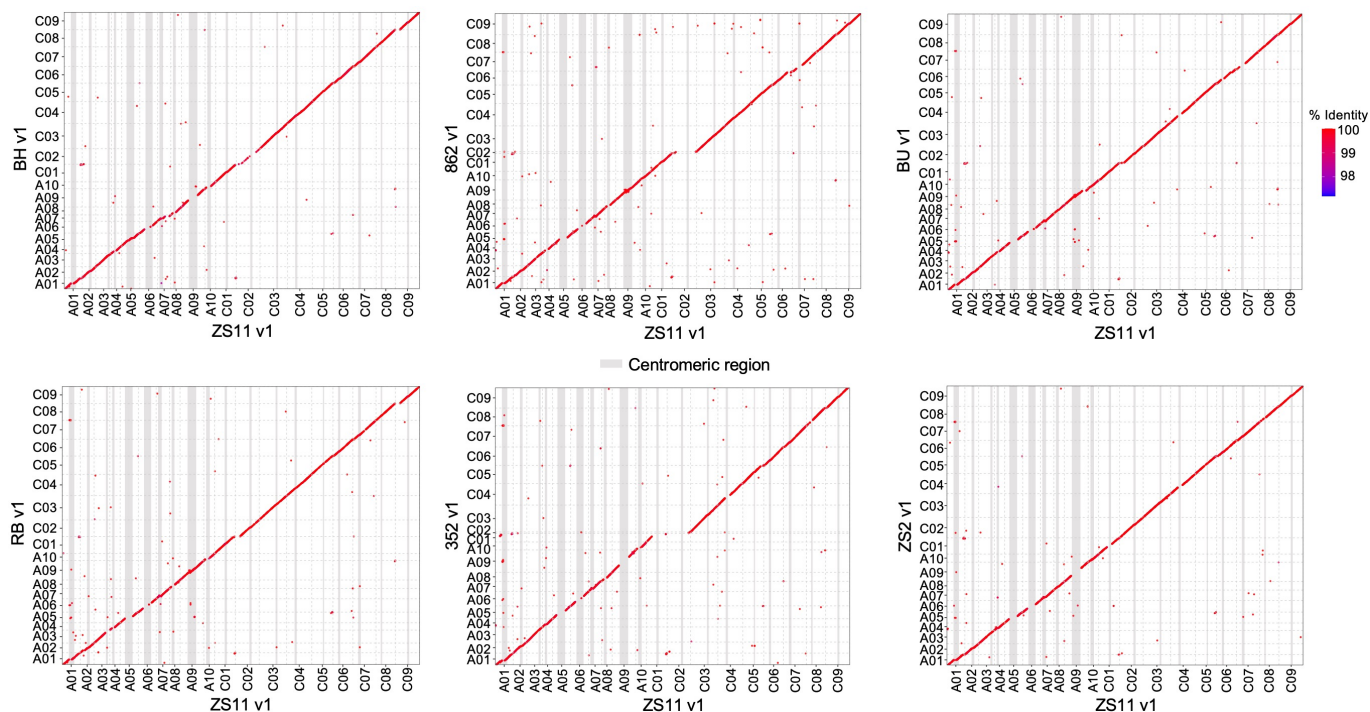
Supplementary Fig 6. Comprehensive validation of 16 *TT4* homologs using long-read mapping. The panels display IGV snapshots for both single-copy genes and tandemly duplicated gene clusters. Green vertical bars highlight the *TT4*.



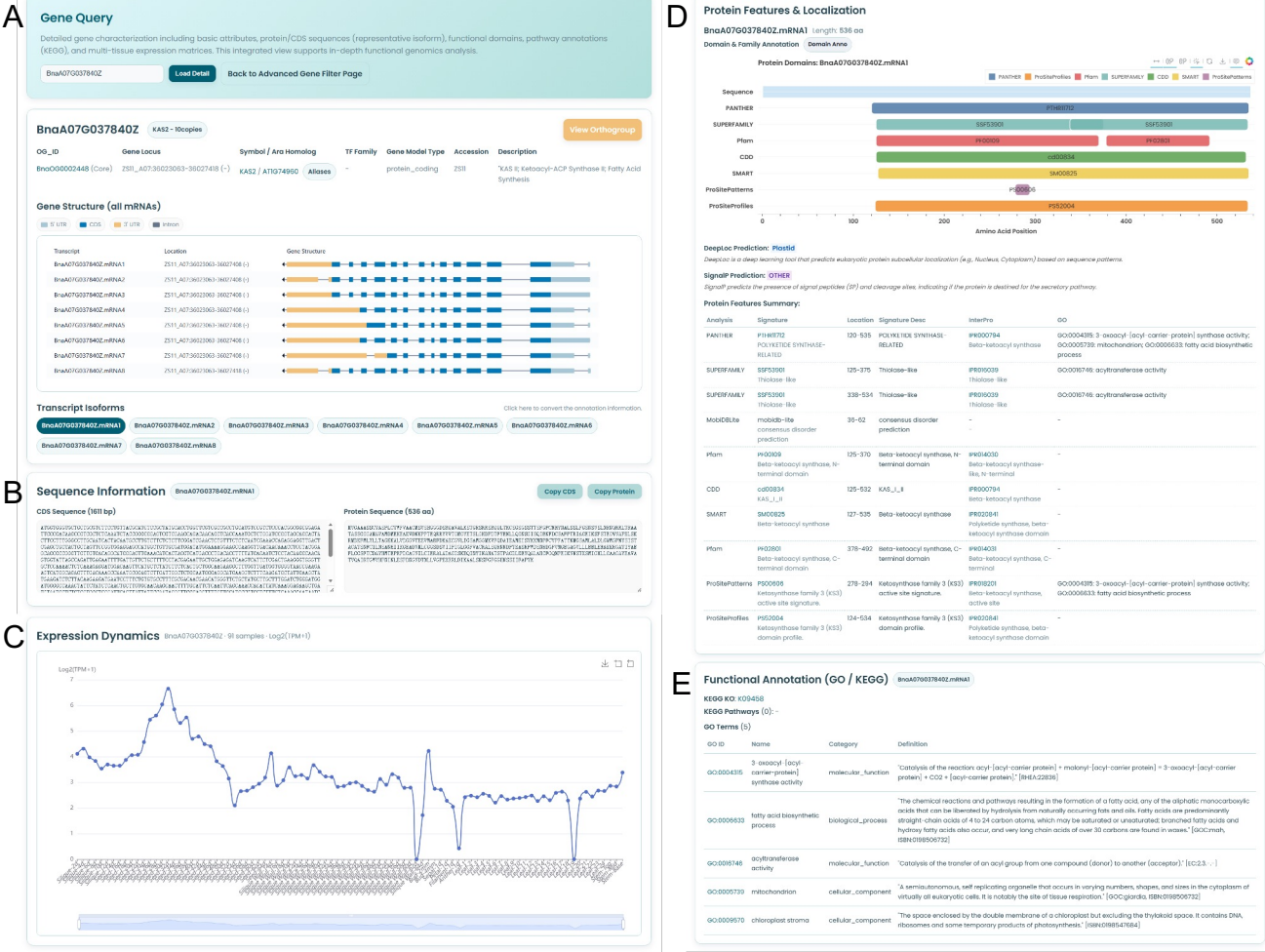
Supplementary Fig 7. Conserved domain analysis of the 16 *TT4* copies.



Supplementary Fig 8. Manhattan plots of GWAS for SGC based on ZS11 v0 genome. The BLUP values of SGC from 2017-2019 were used. The gray dashed line indicates the significance threshold.

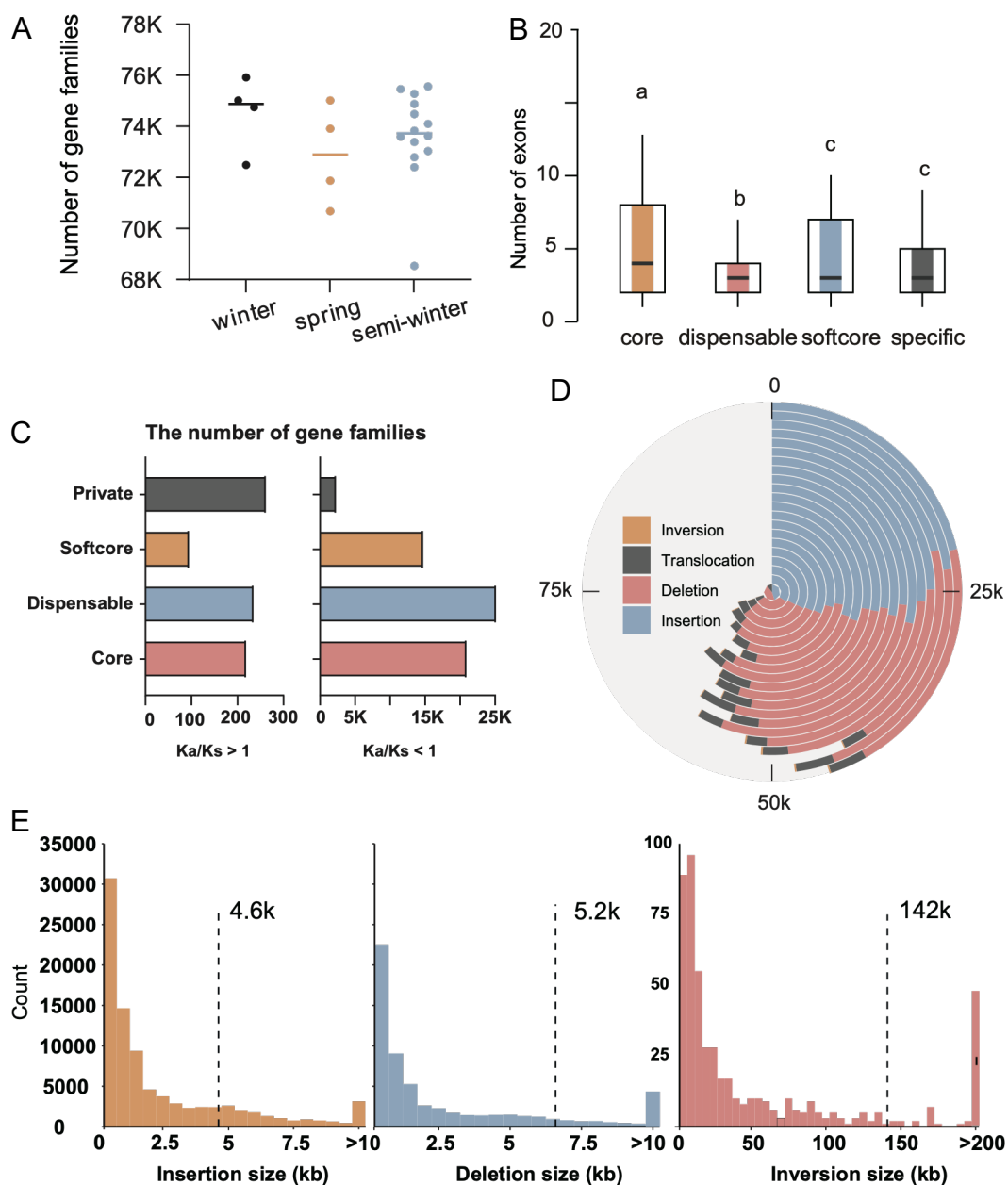


Supplementary Fig 9. Whole-genome synteny dotplots comparing each of the six new assemblies (BH, 862, BU, RB, 352, and ZS2) against the gap-free ZS11 v1 reference genome. The x-axis represents the chromosomes of ZS11 v1, and the y-axis represents the chromosomes of the respective assemblies. The color gradient indicates sequence identity (%), with red denoting high identity. Grey shaded vertical bars indicate the centromeric regions of ZS11 v1.

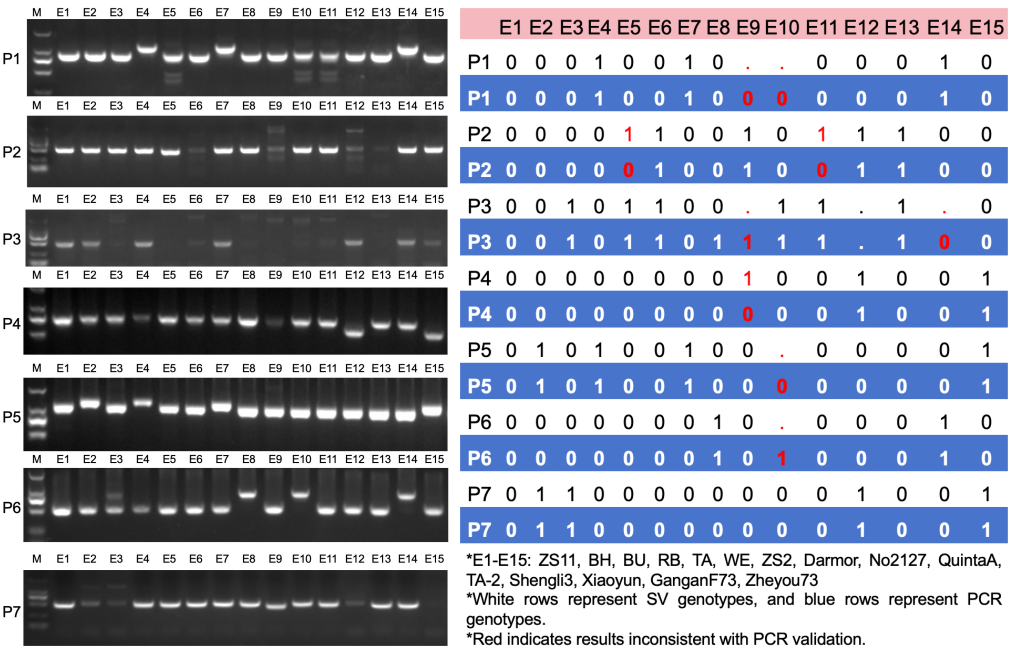




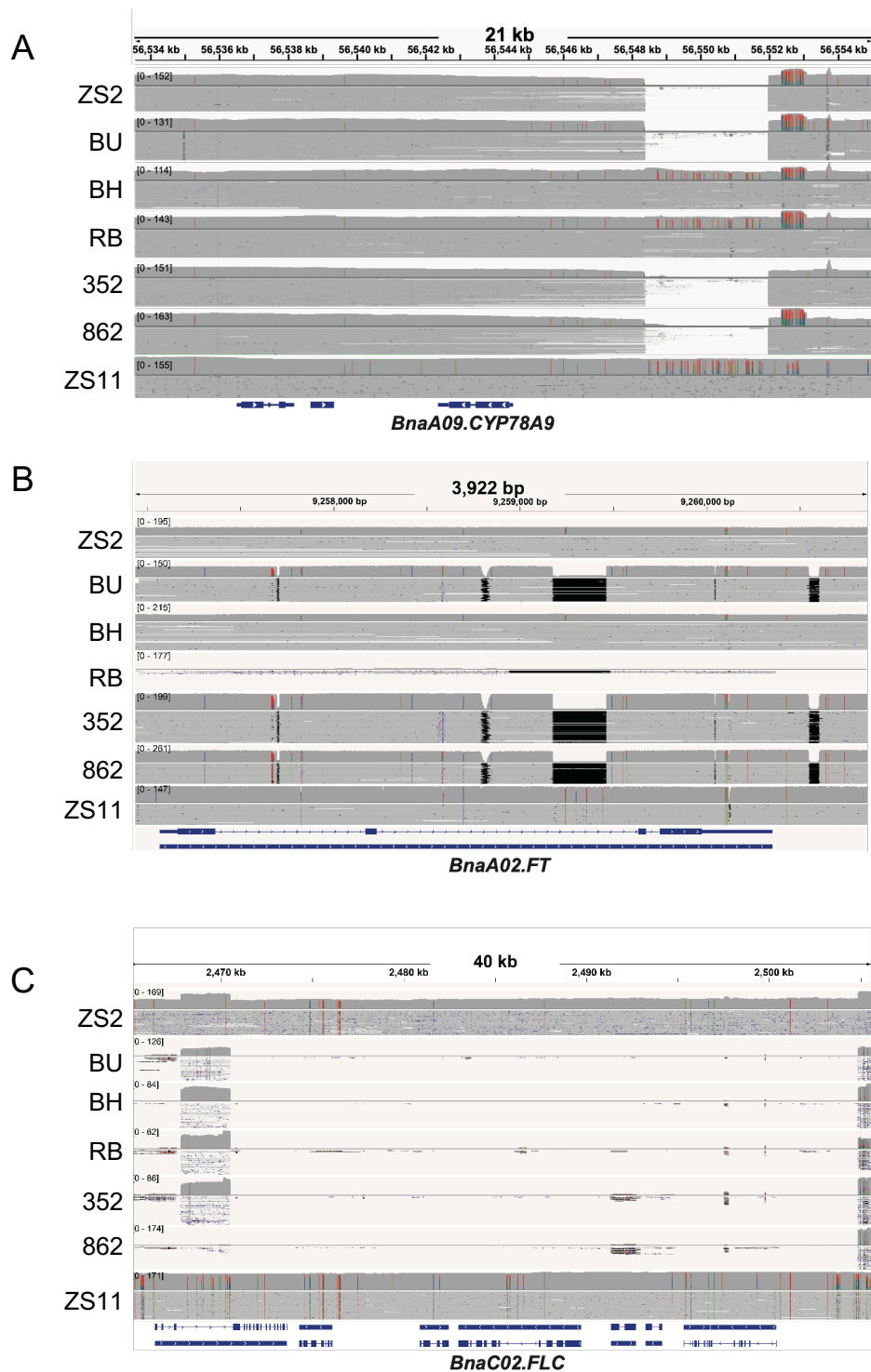
Supplementary Fig 11. Components of the BnaScope Homoeolog Viewer page. **A.** Search tool for functional homoeolog clusters, supporting multiple query methods across 22 genome assemblies. **B.** Basic information of the functional homoeolog cluster and BnaOG assignment. **C.** Phylogenetic distances and gene structures of homoeologs. **D.** Functional domain comparison and multiple sequence alignment (MSA) results of homoeologs. **E.** Expression heatmap of homoeologs.



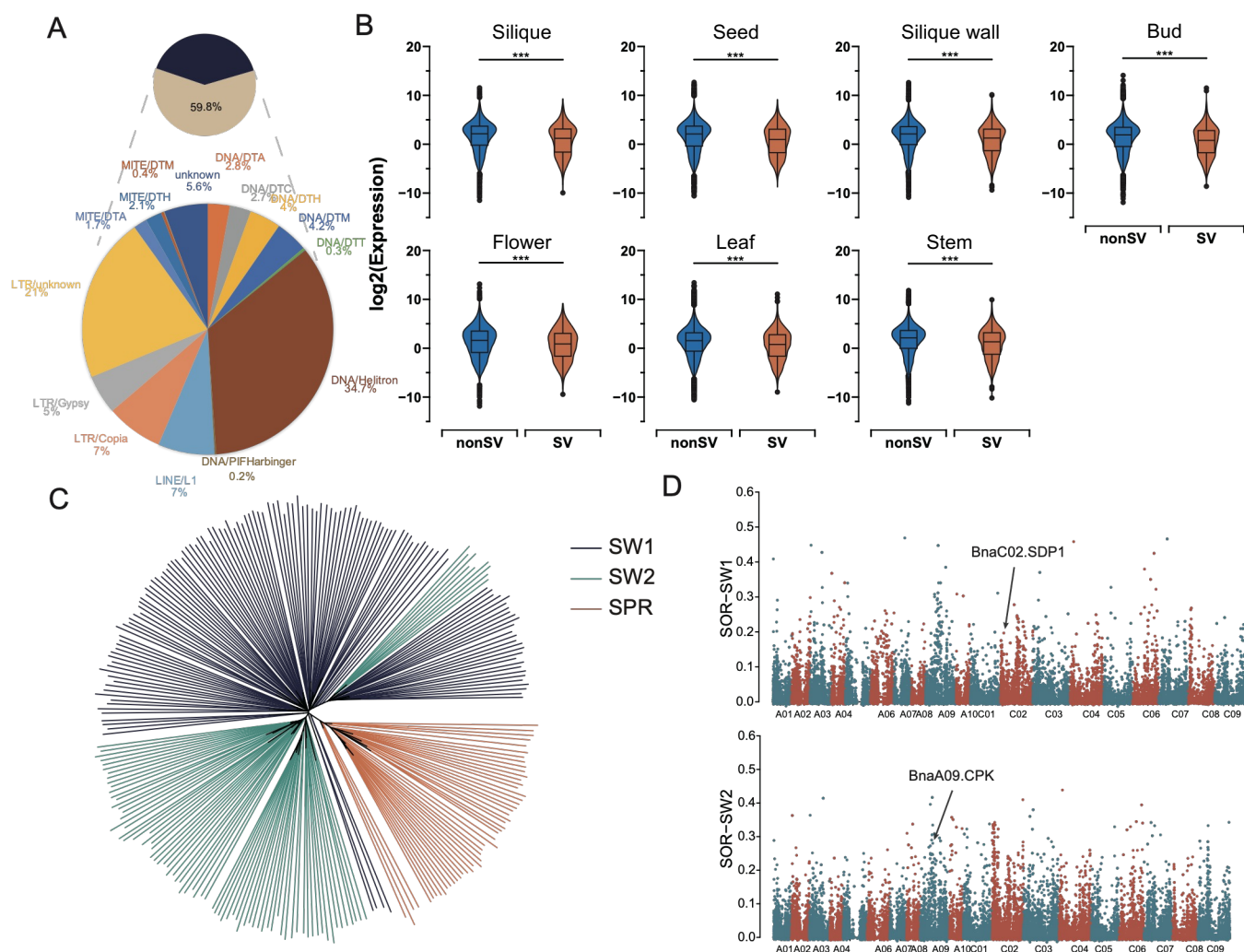
Supplementary Fig 13. Pan-genome and structure variation of *B. napus*. **A.** Gene families in different ecotype genomes in the pan-genome. Sample sizes were: winter, $n = 4$; spring, $n = 4$; semi-winter, $n = 14$. **B.** Number of exons of genes in different clustered gene families. Statistical significance among gene categories was assessed using one-way ANOVA followed by the Student-Newman-Keuls multiple-comparison test with $\alpha = 0.001$. Different letters indicate significant differences between groups. Box plots indicate the median (center line), first and third quartiles (box limits), and values within $1.5\times$ the interquartile range (whiskers). **C.** The number of gene families of different types with average $Ka/Ks > 1$ (or $Ka/Ks < 1$). **D.** Number of different variant types of SV in 22 genomes. **E.** Length of insertions, deletions, and inversion SVs across the genome.



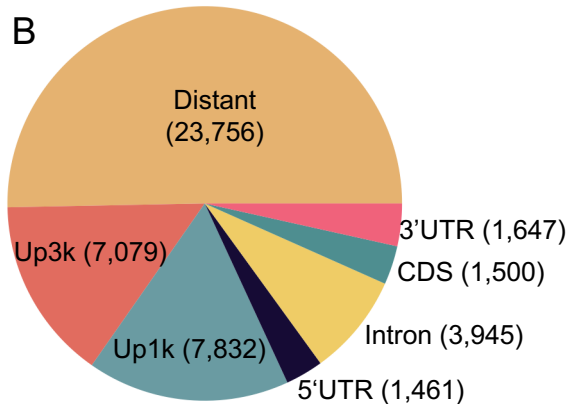
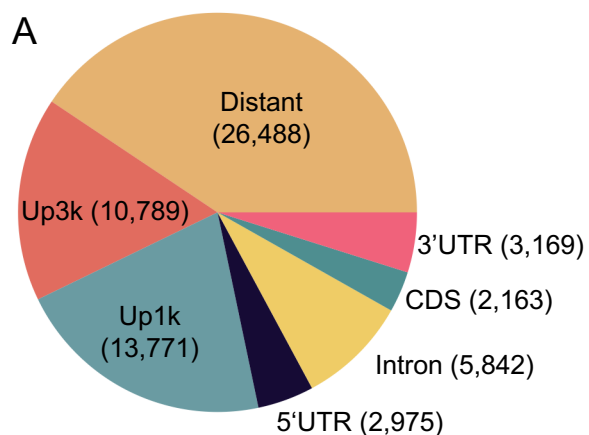
Supplementary Fig 14. PCR validation of seven selected large insertion/deletion SVs across 15 accessions (E1-E15). Left panel: Agarose gel electrophoresis images of PCR amplicons for each primer pair (P1-P7). Right panel: Comparison between predicted SV genotypes (white rows) and PCR-validated genotypes (blue rows). “0” denotes the reference genotype and “1” denotes the alternative genotype. Red numbers indicate discrepancies between predicted and experimental results. The accession names corresponding to codes E1-E15 are listed below the table. Source data are provided as a Source Data file.



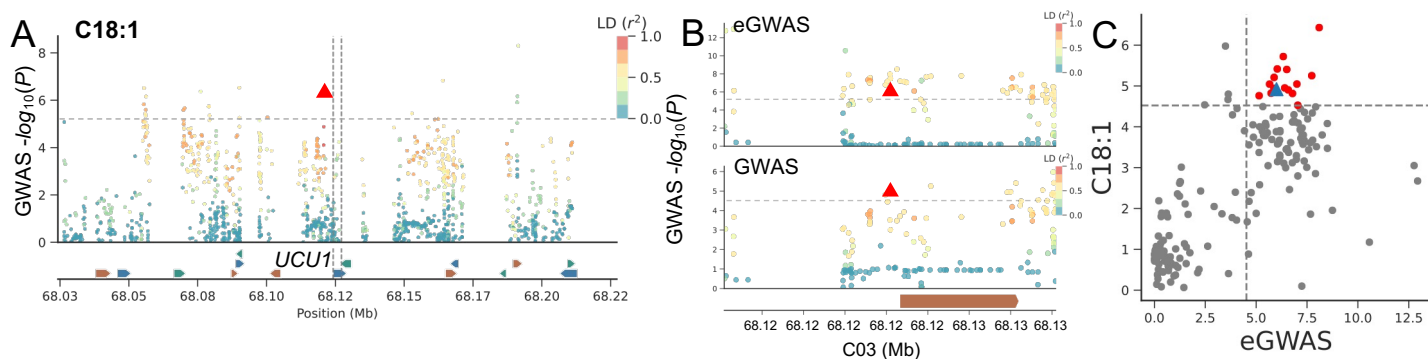
Supplementary Fig 15. Validation of *BnaA09.CYP78A9*(A), *BnaA02.FT*(B) and *BnaC02.FLC*(C) structural variation using raw long reads mapped to the ZS11 genome. **A. A 3.9-kb CACTA-like TE inserted upstream of the *CYP78A9* promoter region in ZS11 (Song *et al.*, 2020). **B.** A 200-bp PAV was found in different ecotypes. **C.** A large deletion including the region of *BnaC02.FLC* was detected in ZS11 and ZS2 ecotypes.**



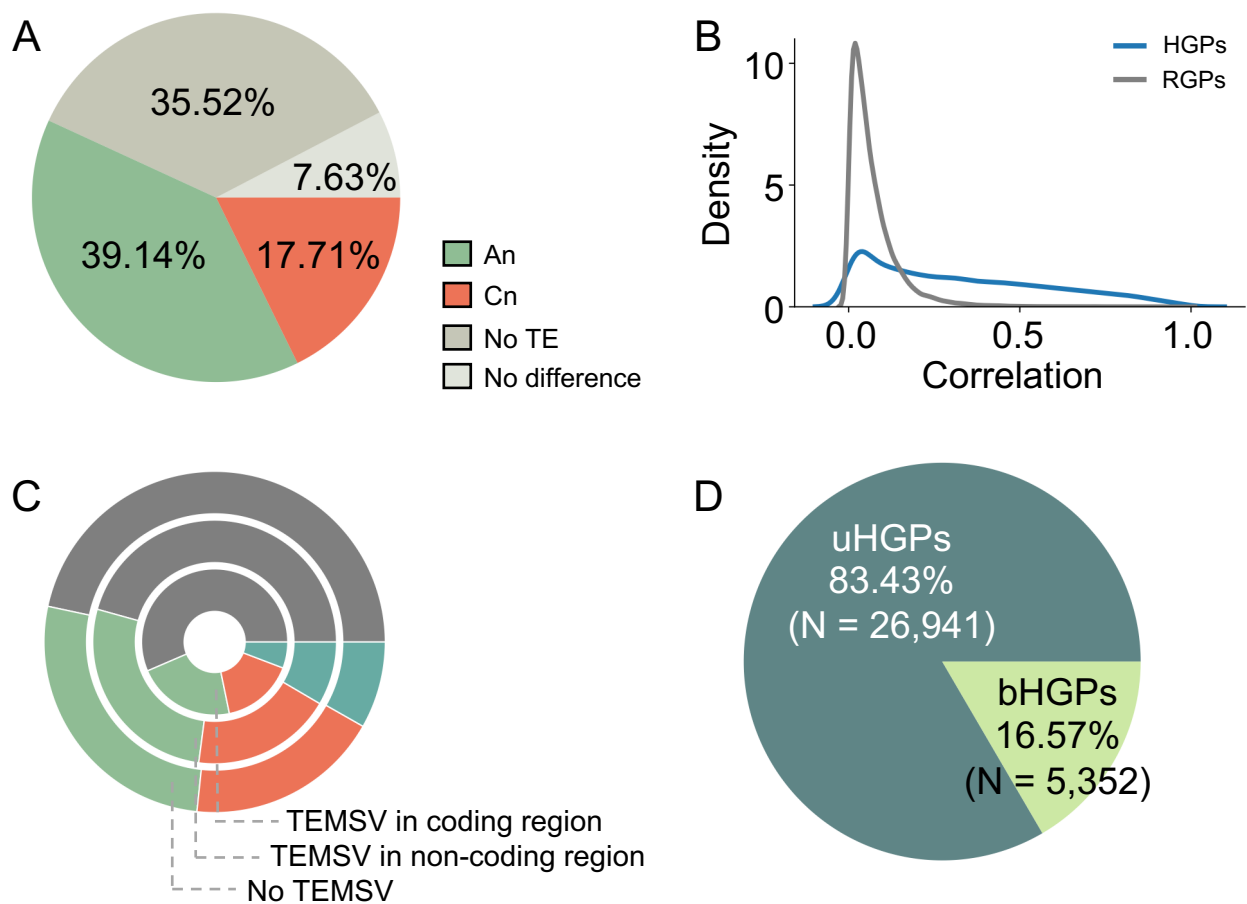
Supplemental Fig 16. The effects of TEMSV on gene expression and the differential distribution of SVs in subpopulations. A. The summary of TE-mediated SVs. **B.** Expression level of SV associated and non-SV gene in different tissues (p-value < 0.001, two-tailed Student's t-test for TPM). n (nonSV) = 107,640; n (SV) = 4,064. Source data are provided as a Source Data file. **C.** Neighbor-joining tree of 505 accessions of three ecotypes were constructed based on SVs. **D.** Genetic differentiation index F_{st} of spring accessions and different semi-winter accessions.



Supplementary Fig 17. Proportion of TEMSVs among subgenomes. (A) is An subgenome, and (B) is Cn subgenome.

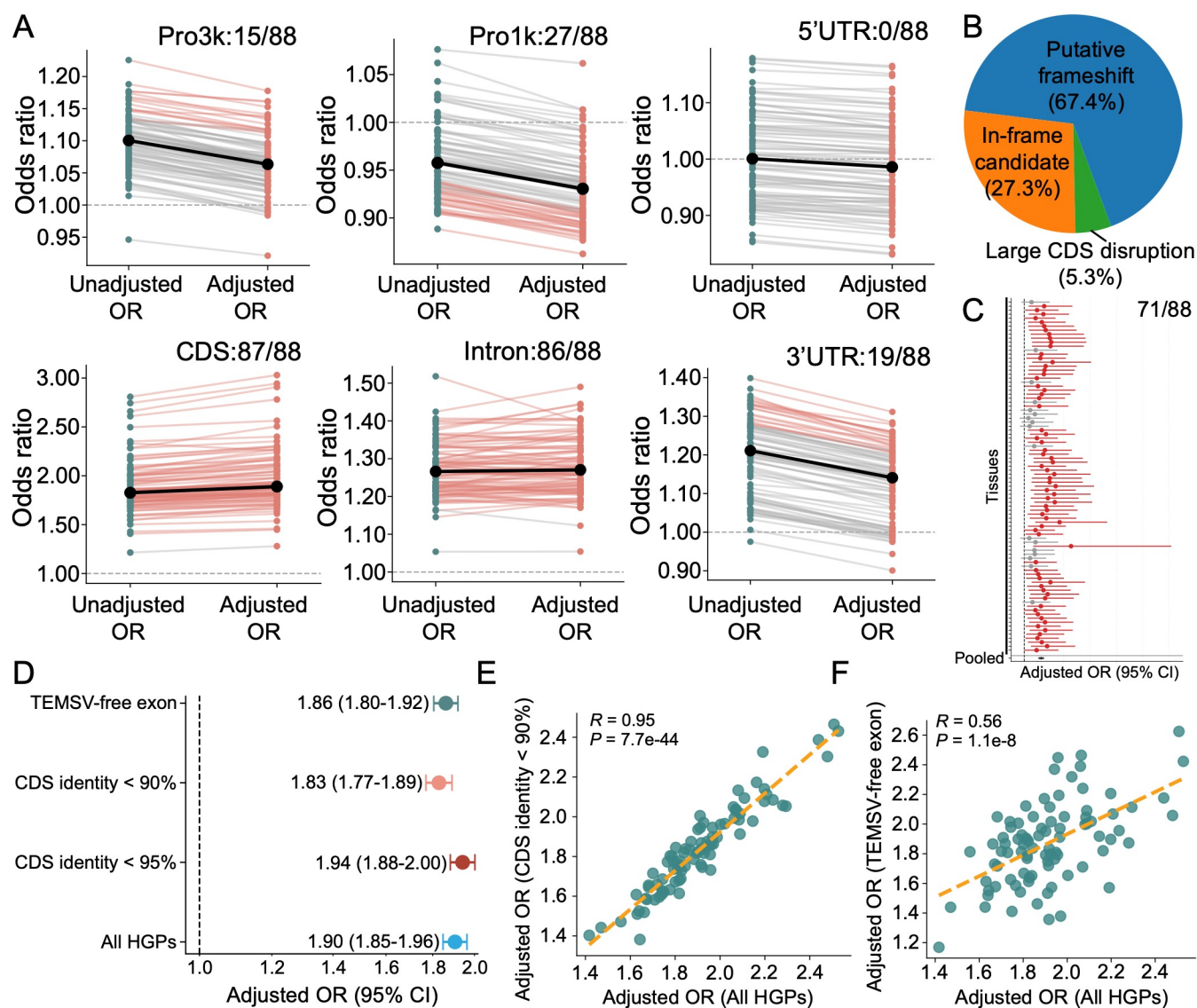


Supplementary Fig 18. *BnaC03.UCU1* is a candidate gene associated with seed C18:1 content. **A.** Manhattan plot of *BnaC03.UCU1* associated loci. The colors on the dots represent the LD between the variant and the Lead SNP, the bottom plot indicates the gene structure in the locus. The gray dashed line indicates the significance threshold. **B.** *BnaC03.UCU1* localized eGWAS as well as GWAS results. Red triangles indicate 592 bp SV. bottom panel indicates the structure of *BnaC03.UCU1*. **C.** Horizontal coordinates indicate log-transformed P values of *BnaC03.UCU1* eGWAS variants in the C03 locus. Vertical coordinates indicate log-transformed P values for the C18:1 GWAS variant in the C03 locus. Blue triangles indicate 592 bp SV.

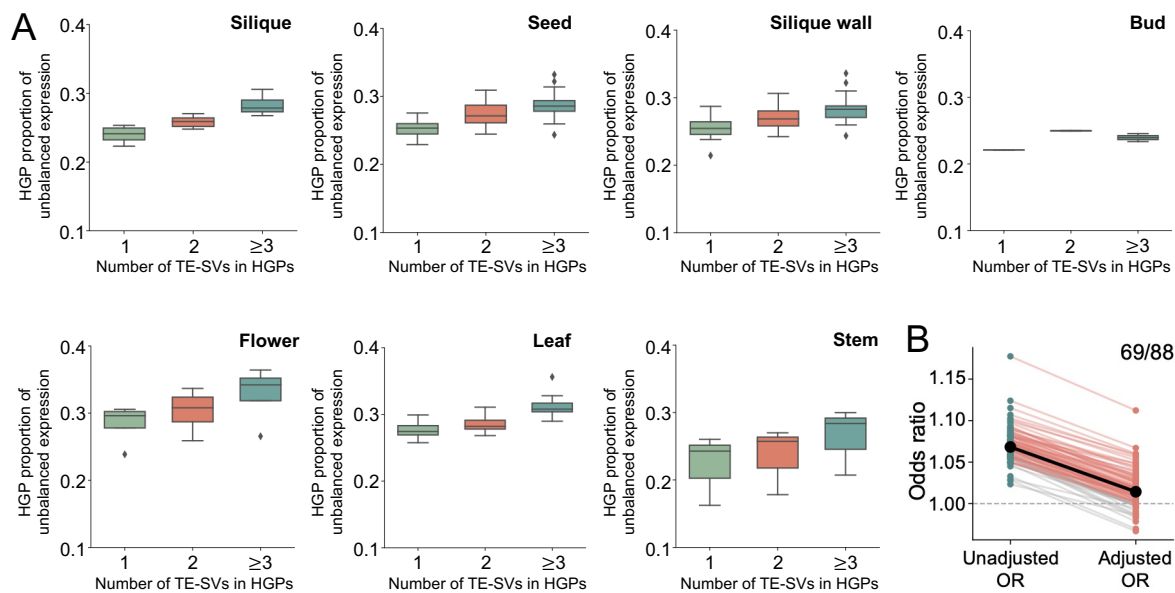


Supplementary Fig 19. Genomic distribution of TEMSVs and characterization of uHGPs.

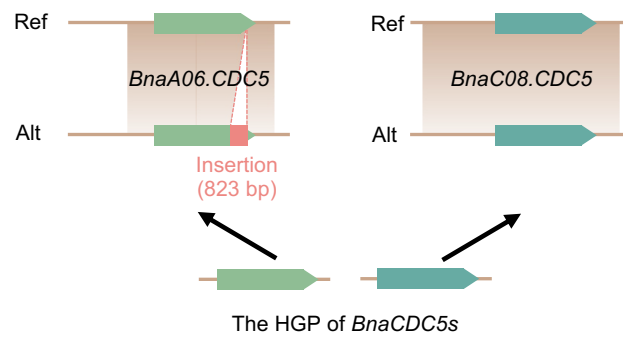
A. Distribution of TEMSV in HGP. Green represents HGPs with a high number of gene TEMSVs in An, red represents HGPs with a high number of gene TE related SVs in Cn, dark gray represents HGPs without TEs, and light gray represents HGPs with no difference in the number of TEs between subgroups. **B.** Correlation distribution of HGP expression in seed at 20 DAF. The blue line indicates the distribution of absolute values of expression correlation between homoeologous copies in HGP. The gray line indicates the correlation distribution performed for 30,784 RGPs. Source data are provided as a Source Data file. **C.** Expression correlation among different HGP types in seeds at 20 DAF. **D.** Proportion of uHGPs in the ZS11 tissues.



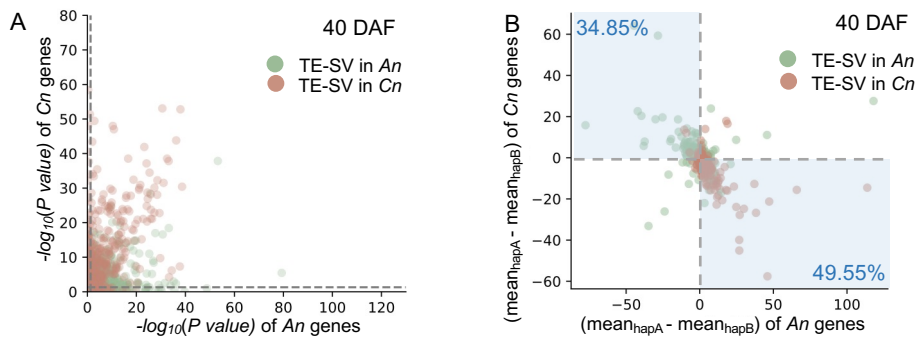
Supplementary Fig 20. Genomic-context adjustment, coding-impact filtering, and expression quantification controls support a robust association between CDS-TEMSVs and uHGP expression. **A.** Comparison of unadjusted and adjusted odds ratios (ORs) across six genomic contexts. For each sample ($n = 88$), the association between TEMSV presence and uHGP status was evaluated using unadjusted OR and adjusted OR. Lines and points are colored coral when the adjusted association remained statistically significant associated with uHGP status ($P < 0.05$), whereas grey indicates non-significant associations. The number of samples satisfying this criterion among 88 comparisons is shown above each panel. The thick black line and enlarged dots indicate the median OR across all 88 comparisons. **B.** Classification of CDS-TEMSVs according to their predicted effects on coding sequences. CDS-TEMSVs were classified into three categories based on their overlap with annotated CDS regions: putative frameshift events, in-frame candidate events, and large CDS-disruption events. **C.** Sensitivity analysis excluding predicted severe coding-disruptive CDS-TEMSVs. Red points and bars indicate significant positive associations ($P < 0.05$ and adjusted OR > 1), whereas grey indicates comparisons that did not meet this criterion. The pooled effect estimate was calculated using a random-effects meta-analysis (DerSimonian–Laird method). **D.** Meta-analysis of CDS-TEMSV associations under different expression quantification and sequence-similarity controls. Pooled adjusted ORs (95% confidence intervals) were calculated across 88 tissue/developmental comparisons for four HGP subsets: all HGPs, HGPs with CDS identity $< 95\%$, HGPs with CDS identity $< 90\%$, and HGPs quantified using TEMSV-free exons. **E.** Correlation between adjusted ORs estimated using all HGPs and the CDS identity $< 90\%$ subset. **F.** Correlation between adjusted ORs estimated using all HGPs and TEMSV-free exon-based expression quantification. Pearson correlation coefficients (R) and significance levels are shown.



Supplementary Fig 21. Association between TEMSV number and unbalanced homoeolog expression after controlling for genomic background effects. A. Effect of the TEMSVs number on the unbalanced expression HGPs in various tissues. The number of HGPs in each category was as follows: 0 TEMSVs, $n = 23,196$; 1 TEMSV, $n = 6,866$; 2 TEMSVs, $n = 1,734$; ≥ 3 TEMSVs, $n = 497$. Box plots indicate the median (center line), first and third quartiles (box limits), and values within $1.5\times$ the interquartile range (whiskers). **B.** Comparison of unadjusted and genomic-context-adjusted odds ratios for the association between HGP-level TEMSV number and uHGP status across 88 tissue or developmental-stage transcriptomic comparisons. An odds ratio greater than 1 indicates a positive direction of association but does not necessarily indicate statistical significance.



Supplementary Fig 22. Structural view of *BnaCDC5s*, showing a 823 bp insertion in *BnaA06.CDC5*.



Supplementary Fig 23. Changes in HGP are affected by TEMSV at 40 DAF. **A.** The degree of change in HGP expression at 40 DAF. Green dots indicate TEMSV located near the *An* copy of HGP. Red dots indicate TEMSV located near the *Cn* copy of HGP. The horizontal coordinate represents the P values for expression differences between *An* genes in HGPs with and without TEMSV, calculated by a two-sided Student's t -test. The vertical coordinate represents the P values for expression differences between *Cn* genes in HGPs with and without TEMSV, calculated by a two-sided Student's t -test. Source data are provided as a Source Data file. **B.** The direction of change in HGP expression at 40 DAF. Green dots indicate TEMSV located near the *An* copy of HGP. Red dots indicate TEMSV located near the *Cn* copy of HGP. The horizontal coordinate indicates the difference between the expression means in the HGP between *An* genes in HGPs with and without TEMSV. The vertical coordinate indicates the difference between the expression means in the HGP between *Cn* genes in HGPs with and without TEMSV. Shading indicates that homoeologous genes in HGP change in the opposite direction. Source data are provided as a Source Data file.

Supplementary File: Bnascope - Feature Introduction

Introduction

Brassica napus (*B. napus*, AACCC, $2n=38$) is a typical allotetraploid crop whose genome has experienced complex whole-genome duplication events and rearrangements between homoeologous chromosomes¹. As a vital oilseed crop, the genomic complexity of *B. napus* leads to high gene redundancy and dosage effects, making accurate differentiation and functional verification of homoeologs extremely challenging, thereby impeding the identification of agronomic genes². Previous studies have also revealed imbalances between subgenomes in terms of gene expression, epigenetic activity, and single nucleotide polymorphism (SNP) variations^{3, 4}. Although the rapeseed research community is highly active and has accumulated vast amounts of omics data—establishing databases such as BnIR, BRAD, and BrassicaEDB—existing platforms tend to focus on omics data visualization and lack a comprehensive system that integrates homoeologous relationships and multi-omics evidence with a "gene-centric" dimension^{5, 6, 7}.

Based on the 22 genome assemblies constructed in this study, we established the BnaScope database to provide a detailed and rich gene functional interface. It aims to assist researchers in dissecting functional divergence between subgenomes and efficiently identification key agronomic genes. As a unified community reference resource, BnaScope integrates various high-precision prediction tools and multi-omics data, offering standardized gene functional reports and an interactive homoeolog analysis interface (Table S1). The platform is freely open to the scientific community (<https://rgmi.hzau.edu.cn/bnascope/>) without registration and will be continuously updated to incorporate the latest omics resources and refined functionalities.

Unified Gene Reporting Page

To address inconsistent annotation standards across multiple genome versions, BnaScope implemented a standardized gene nomenclature

(Bna[Chr]G[Number][Suffix]) for the 22 collected reference genomes. We constructed a standardized gene reporting page that integrates basic genomic attributes with deep functional predictions ([Supplementary Fig. 8](#)). For common variety assemblies such as ZS11, Westar, and Darmor, regarding functional annotation, we systematically annotated 463,895 protein sequences using InterProScan v5.76-107.0, covering multiple databases including Pfam, PANTHER, CDD, ProSiteProfiles, ProSitePatterns, SUPERFAMILY, PRINTS, and MobiDBLite, alongside GO term assignment^{8, 9}. Additionally, subcellular localization was predicted using the protein language model-based DeepLoc v2.1, signal peptides were predicted using SignalP v6.0, and KEGG Orthology (KO) identifiers were mapped via KofamScan^{10, 11, 12}. Subsequently, we aligned 273 RNA-seq datasets (covering 91 tissues including roots, stems, leaves, flowers, seeds, and siliques) to the ZS11 reference genome using Salmon to construct tissue-specific expression patterns, providing direct transcriptomic evidence for gene function inference¹³.

Homoeolog Viewer

Functional redundancy and divergence among homoeologous gene copies represent a critical challenge in rapeseed research. BnaScope has constructed a specialized homoeolog analysis module ([Supplementary Fig. 9](#)). We first established homoeologous gene clusters (termed BnaOG) across the 22 genomes based on OrthoFinder, achieving cross-variety clustering of homoeologs¹⁴. Furthermore, by BLAST aligning protein sequences from all genomes to the Arabidopsis Araport11 database, we accurately mapped Arabidopsis functional annotations to the rapeseed genomes^{15, 16}. All rapeseed copies aligning to the same Arabidopsis gene (e.g., *FLC*) were defined as functional homoeologs.

Leveraging Bokeh visualization components, we developed an interactive Homoeolog Viewer. Users need only input a unified gene Symbol to simultaneously retrieve and display all copies of that gene across the A/C subgenomes. This interface integrates multi-dimensional visualization views,

including phylogenetic trees (genetic distance), gene structure comparisons, expression heatmaps (analysis of homoeolog expression bias), and protein multiple sequence alignments (MSA) powered by MUSCLEv5 with domain distribution¹⁷. Collectively, these tools constitute a key chain of evidence for distinguishing the functional specificity of homoeologous copies.

Multi-omics Data Integration

BnaScope deploys a high-performance JBrowse2 genome browser, enabling the integrated display of genomic structural features and population variation information¹⁸ ([Supplementary Fig. 10](#)). Within the genome tracks, we integrated 1,058,755 transposable elements (TEs) annotated by the EDTA pipeline, as well as SNP/InDel variation information derived from 505 resequencing accessions^{19, 20}. We anticipate that these data will facilitate user exploration of TEMSVs and their potential regulatory effects. Additionally, for the reference genome ZS11 v1, the platform re-analyzed and aggregated multi-omics data including RNA-seq, ATAC-seq, and ChIP-seq^{3, 13, 21}. Users can synchronously view chromatin accessibility, histone modifications, and gene expression signals in the browser, thereby dissecting cis-regulatory mechanisms of gene expression. Detailed parameters and methods regarding data processing can be found on the website's Data Notes page.

Toolkits

BnaScope provides a comprehensive online analysis toolkit. It offers a highly integrated search tool supporting multi-dimensional retrieval based on Gene ID, keywords, chromosomal location, Symbol, functional domain (PFAM), gene family (PANTHER), and GO terms. Meanwhile, the platform incorporates a high-performance BLAST+ service, supporting comprehensive homology searches against 26 genome versions (including 22 new assemblies and 4 classic versions)^{1, 16, 22, 23}. Notably, to address data compatibility issues caused by version updates, we developed an ID Conversion tool that supports bidirectional conversion between old and new version IDs. Furthermore, the platform provides Flanking Sequence Fetch, enrichment analysis, and

comprehensive data download services, maximizing support for personalized data mining requirements.

Table S1. Feature Overview of BnaScope Database

Category	Functionality	Assembly Coverage
Gene Annotation	Standardized Gene Nomenclature	22 Assemblies
	Protein Domain Annotation (InterProScan)	ZS11 v1 / Westar v1 / Darmor v1
	Gene Ontology (GO) Annotation	ZS11 v1 / Westar v1 / Darmor v1
	KEGG Pathway Annotation	ZS11 v1 / Westar v1 / Darmor v1
	Subcellular Localization (DeepLoc 2.1)	ZS11 v1 / Westar v1 / Darmor v1
	Signal Peptide Prediction (SignalP 6.0)	ZS11 v1 / Westar v1 / Darmor v1
Homoeolog Analysis	Homoeolog Cluster Definition	22 Assemblies
	Functional Homoeolog Mapping (via <i>Arabidopsis</i>)	22 Assemblies
	Homoeolog Comparative View	22 Assemblies
	Domain Alignment & Expression Profiling	22 Assemblies
JBrowse2 Multi-omics	Genome & Gene Model Tracks	22 Assemblies
	TE Annotation Tracks	22 Assemblies
	Population Variation Tracks	ZS11 v1
	ChIP-seq Tracks	ZS11 v1
	ATAC-seq Tracks	ZS11 v1
	RNA-seq Tracks	ZS11 v1
Toolkits	Gene & Annotation Search	22 Assemblies
	Flanking Sequence Retrieval	22 Assemblies
	BLAST+ Server	26 Assemblies (22 New + 4 Legacy)
	Gene ID Conversion	26 Assemblies (22 New + 4 Legacy)
	Enrichment Analysis (GO / KEGG)	ZS11 v1 / Westar v1 / Darmor v1
	Data Download	22 Assemblies

Supplementary References

1. Chalhoub B, *et al.* Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *science* **345**, 950-953 (2014).
2. Gu J, Guan Z, Jiao Y, Liu K, Hong D. The story of a decade: Genomics, functional genomics, and molecular breeding in *Brassica napus*. *Plant Communications* **5**, (2024).
3. Zhang Q, *et al.* Asymmetric epigenome maps of subgenomes reveal imbalanced transcription and distinct evolutionary trends in *Brassica napus*. *Molecular Plant* **14**, 604-619 (2021).
4. Lu K, *et al.* Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature communications* **10**, 1154 (2019).
5. Yang Z, *et al.* BnLR: A multi-omics database with various tools for *Brassica napus* research and breeding. *Molecular Plant* **16**, 775-789 (2023).
6. Chen H, *et al.* BRAD V3. 0: an upgraded Brassicaceae database. *Nucleic Acids Research* **50**, D1432-D1441 (2022).
7. Chao H, *et al.* BrassicaEDB: a gene expression database for Brassica crops. *International journal of molecular sciences* **21**, 5831 (2020).
8. Blum M, *et al.* InterPro: the protein sequence classification resource in 2025<? mode longmeta?>. *Nucleic acids research* **53**, D444-D456 (2025).
9. Jones P, *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
10. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252 (2020).
11. Teufel F, *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature biotechnology* **40**, 1023-1025 (2022).
12. Ødum MT, Teufel F, Thumhuri V, Almagro Armenteros JJ, Johansen AR, Winther O, Nielsen H. DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Research* **52**, W215-W220 (2024).
13. Tan Z, *et al.* Comprehensive transcriptional variability analysis reveals gene networks regulating seed oil content of *Brassica napus*. *Genome biology* **23**, 233 (2022).
14. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**, 238 (2019).
15. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal* **89**, 789-804 (2017).
16. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC bioinformatics*

- 10**, 421 (2009).
17. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature communications* **13**, 6968 (2022).
 18. Diesh C, *et al.* JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome biology* **24**, 74 (2023).
 19. Tang S, *et al.* Genome-and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Molecular Plant* **14**, 470-487 (2021).
 20. Ou S, *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology* **20**, 275 (2019).
 21. Hu J, *et al.* Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. *Nature Genetics* **54**, 694-704 (2022).
 22. Song J-M, *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature plants* **6**, 34-45 (2020).
 23. Rousseau-Gueutin M, *et al.* Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience* **9**, giaa137 (2020).