Genome assembly and genomic architecture of a prominent cold-resistant rapeseed germplasm

Zefeng Wu[1], Guoqiang Zheng[1], Yali Sun[1], Xiaoyun Dong[1], Ying Wang[1], Hui Li[1], Yinin Niu[1], Zhen Huang[2], Jihong Hu[2], Zigang Liu[1*]

[1]State Key Laboratory of Aridland Crop Science, Gansu Agricultural University, Lanzhou, 730070, Gansu, China

[2]State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, 712100, Shaanxi, China.

[*]Corresponding author: lzgworking@163.com

Word count

Summary: 167

Introduction: 720

Results: 3916

Discussion: 696

Materials and methods: 2896

Abstract

Cold stress is one of the most limiting factors affecting crop production. Winter rapeseed, as an important oilseed crop, is often subjected to low temperature stress and consequent yield loss at high latitudes. Therefore, identification of highly cold-resistant winter type rapeseed and elucidation of the molecular mechanism of cold tolerance are crucial for genetic improvement of the trait. Here, we report a chromosome-level reference genome of a prominent cold-resistant winter rapeseed cultivar, NTS57. We elucidated the phylogeny of NTS57 with other public rapeseed cultivars and performed gene family and structure variation analysis at the pan-genome level. In addition, we profiled the DNA methylation and transcriptomic changes of NTS57 upon cold stress, and many cold response genes were identified. Furthermore, we surveyed the cold resistance of 290 rapeseed accessions and performed GWAS analysis to identify the potential gene loci controlling cold resistance, which discovered some important functional genes, including *MYB34* and *RLK1.* Overall, this research provided a basis for genetic dissection and improvement of cold tolerance in rapeseed.

Keywords

Cold stress; rapeseed; genome assembly; RNA-Seq; methylation; GWAS

Introduction

Cold stress, including chilling (0-15℃) and freezing (< 0℃), is one of environmental factors affecting plant growth, productivity, and geographical distributions of species (Ding et al., 2019). Most temperate plants have evolved the ability to adapt to low temperatures, referred to as cold acclimation, and show improved freezing tolerance after exposure to non-freezing or non-lethal low temperatures (Browse and Xin, 2001). During cold acclimation or chilling stress, plants undergo a number of physiological and biochemical changes, including transcriptional, translational, and metabolic changes that increase or decrease the levels of specific proteins, metabolites, and phytohormones (Ding et al., 2019; Guo et al., 2018; Kidokoro et al., 2022). First, cold stress affects membrane rigidification, which is one of the most prominent

physiological indicators and further triggers downstream cold stress responses in plants (Orvar et al., 2000). This situation results in the extravasation of cell sap and an increase in relative electrical conductivity (REC) (Lu et al., 2017). Thus, REC has become a commonly used index to assess membrane permeability and measure the effect of low temperature on plants (Cui et al., 2020; Guan et al., 2019; Yu et al., 2021). Second, chilling stress disrupts the stability of proteins or protein complexes and reduces the activities of enzymes such as ROS-scavenging enzymes (Ding et al., 2019; Yu et al., 2020). These physiological changes may be mirrored by biochemical responses. For example, many up-regulated genes during cold acclimation are involved in the synthesis of protective substances, such as soluble sugars and proline, and proteins including late embryogenesis abundant (LEA) proteins, anti-freezing proteins (AFP), and cold shock proteins (CSPs), while the down-regulated genes have functions in photosynthesis, such as genes encoding photosystem I subunits (*Psa*) and photosystem II subunits (*Psb*) (Kidokoro et al., 2022). Although much progress has been made in the last two decades in identifying the crucial molecular components involved in cold stress tolerance and in dissecting their regulatory mechanisms, we are still far from a good understanding of the molecular mechanism underlying cold signal perception and transduction in plants, especially for important economic crops.

Rapeseed (*Brassica napus* L.) is one of the most important economic crops in the world. There are three major ecotypes of rapeseed, including spring, winter, and semi-winter types, which are the result of evolution and long-term cultivation to adapt to different climates (Wei et al., 2017). Among them, winter-type rapeseed is one of the few crops that can survive frosty winters in temperate regions (Liu et al., 2022). The Yangtze River basin is the most important ecological region for rapeseed production in China. Winter- and semi-winter rapeseed are cultivated in the Yangtze River basin, with seeds sown from late September to mid-October, which means that the growth and production of rapeseed are always threatened by the cold temperature (Ke et al., 2020). More importantly, in northern China, where winters are extremely cold, most winter rapeseed cultivars could not successfully overwinter until some highly cold-tolerant rapeseeds were bred. “NTS57” is currently the most cold-tolerant variety of winter

rapeseed bred by our research group, as it can survive temperatures as low as -26 ℃ with an overwintering rate of over 90% without snow cover (Mi et al., 2021; Wei et al., 2021). Unraveling the molecular mechanisms of these highly cold-tolerant rapeseeds would not only help us better understand the cold stress adaptation of plants, but is also of great agricultural importance for the northward expansion of winter rapeseed in China and other countries.

Currently, the genomes of Darmor-*bzh* and ZS11, representatives of two different ecotypes of rapeseed, have been assembled and serve as the excellent reference for functional genomics study of rapeseed (Chalhoub et al., 2014; Chen et al., 2021; Sun et al., 2017). More recently, the germplasm of eight rapeseed varieties has also been de novo assembled, providing a pan-genomic view and a valuable genomic resource for rapeseed breeding (Song et al., 2020). However, none of these varieties has been recognized as a prominent cold-resistant germplasm, which hinders the improvement and promotion of cold-resistant varieties in cold regions. In the present study, we assembled a chromosome-level reference genome of winter rapeseed, NTS57. The genome assembly enabled us to explore the molecular mechanism of cold resistance through pan-genome comparison, phylogenetic, epigenomic, transcriptomic and GWAS analysis. Our work provides a high-confidence framework for studying and breeding crops with strong cold tolerance.

Results

Genome assembly and annotation of a cold tolerant rapeseed NTS57

B. napus NTS57 (2n=19, AACC) is a cultivated rapeseed developed by crossing winter rapeseed with *Brassica rapa*. It has strong cold resistance that can survive in the field at -26 ℃ in northwest China. For example, NTS57 plants grew and developed well under normal conditions, similar to ZS11 and Westar (Figure 1a). However, leaves of the ZS11 and Westar accessions began to wilt after 12 hours at -4 ℃, while NTS57 appeared normal and healthy (Figure 1a). At 24 h after treatment, NTS57 was still growing better than the other two accessions, and only some of its leaves were curled or wilted (Figure 1a). When all plants were transferred to normal temperature and left

109 for a period of time, NTS57 could fully recover its phenotype, while the other two
110 accessions could not or only partially recover (Figure 1a). Apparently, NTS57 can
111 survive cold stress and will be an important germplasm for the cold tolerance research
112 in rapeseed.

113 To explore the genetic architecture contributing to the cold resistance trait of
114 NTS57, we performed DNA sequencing and de novo assembly of the NTS57 genome
115 by integrating 60x PacBio high-fidelity (HiFi) long-read sequencing, 100x Hi-C (high-
116 throughput chromosome conformation capture) sequencing, and 30x Illumina paired-
117 end sequencing data. First, K-mer analysis of the Illumina sequencing reads estimated
118 the genome size of NTS57 to be *c.* 1.01~1.29 Gb (Figure S1). The PacBio reads were
119 then assembled and polished into a 1.06 Gb genome assembly with contig N50 21.81
120 Mb, and more than 93.6% (995 Mb) of the sequences were anchored on 19
121 chromosomes with scaffold N50 58.64 Mb after Hi-C scaffolding (Figure 1b and c).
122 Centromere sequences were successfully identified for all 19 chromosomes of NTS57
123 (Figure 1b and c; Table S1). The high completeness of the assembly was also evidenced
124 by a BUSCO recovery score of 99.75%. Accordingly, the assembly quality of the
125 present sequenced genome is higher than the published rapeseed reference genomes
126 (Table 1). Based on ab initio prediction, homology search and transcriptome mapping
127 strategies, we obtained a total of 101,438 protein-coding genes in the NTS57 genome.
128 Of the gene set, 96,922 (95.5%) were located on 19 chromosomes, with 44,411 in the
129 A subgenome and 52,511 in the C subgenome, which was close to the gene number in
130 Darmor-*bzh* and ZS11 genomes (Table S2). Among them, we identified 24,217
131 homologous gene pairs based on reciprocal blast search, and the gene expression
132 analysis revealed that 6,994 (28.9%) gene pairs on the homologous chromosomes
133 showed at least a twofold gene expression difference (Figure S2). Although there were
134 936 more gene pairs where the C-homeolog was more highly expressed than the A-
135 homeolog, no evidence for remarkable subgenome dominance was observed
136 considering the overall strong gene expression correlations (PCC = 0.81; P-value <
137 2.2e-16) between two subgenomes (Figure S2).

138 Of the 101,438 predicted genes, more than 78% were supported by transcriptomic

data. By homology search, 97.89% were well annotated by at least one of the following protein-related databases: NCBI nr database, the Swiss-Prot protein database, the KOG database, KEGG database, and the GO database (Figure S3). In addition, 36,552 non-coding RNAs, including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), microRNAs (miRNAs) and small nuclear RNAs (snRNAs), were annotated in the NTS57 genome (Table S3). These results indicated the high quality of the NTS57 genome assembly.

The amplification of various repetitive DNA elements is one of the key drivers of plants evolution. Annotation of the NTS57 genome revealed that more than 55.6% of the genomic content was repetitive elements (Table S4). Among these DNA repeats, 80% were TEs, including retrotransposons and DNA transposons. The Gypsy and Copia represented the highest proportion of the retrotransposons, accounting for 8.6% and 5.1% of the whole genome, respectively (Table S4). Genomic visualization analysis revealed that the two types of transposons had overall similar distribution patterns along the chromosomes and an anti-correlation with gene density (Figure. 1b). However, we noticed that they displayed an opposite pattern at some local positions, especially near centromeres with high GC content and DNA methylation level (Figure 1b), which is similar to a previous study on the ZS11 genome (Chen et al., 2021). In addition, chromosomal synteny analysis based on protein-coding gene arrangements revealed that there is strong intraspecific collinearity between homologous chromosomes from two subgenomes (Figure 1b), further demonstrating the homology of chromosomes inherited from two ancestral diploid genomes.

Phylogeny and pan-genomic analysis of rapeseed varieties

To reveal the phylogeny of NTS57 with other rapeseed cultivars, we integrated an additional 10 publicly sequenced rapeseeds with different eco-types, including three winter-type (Darmor-*bzh*, Quinta, and Tapidor), four semi-winter-type (ZS11, Shengli, Zheyou, and Gangan), and two spring-type (Westar and No2127) rapeseed cultivars. To reduce the clustering bias of the orthologous genes derived from two subgenomes, all the investigated rapeseed genomes were split into A and C subgenomes, resulting in 22

subgenomes for further analysis. In addition, Arabidopsis and two diploid progenitors, *Brassica rapa* and *Brassica oleracea*, were added as outgroup species to infer the species tree. The max-likelihood species tree inferred based on 2,268 single-copy orthologous genes revealed that the A subgenome of NTS57 were closed to that of semi-winter rapeseed, while the C subgenome formed a clade with that of winter-type rapeseed (Figure 2a). This result indicated that the NTS57 and semi-winter rapeseed bred in China had a different breeding history, and the cold tolerance may inherit or kept from winter type rapeseeds.

While research on pan-genomes for rapeseed has been reported (Song et al., 2020), there is currently limited research on the cold tolerance of rapeseed at the pan-genome level. To this end, we investigated the pan-genome architecture by integrating NTS57 and 10 other rapeseed accessions. We identified a total of 43,734 pangene families by clustering protein-coding genes from the A sub-genomes and 54,091 from the C sub-genomes. The number of gene families in both A subgenome and C subgenome increased rapidly when more subgenomes or species were added (Figure 2b), suggesting that the 11 rapeseed genomes have rich gene pools and that a single reference genome cannot capture the full genetic diversity of rapeseed. For the A subgenome, more than 18,375 gene families were conserved among the 11 rapeseed accessions, representing the core gene families. The dispensable gene families that present in two to 10 accessions occupied 55.69% of the pan-gene families, and 2.3% of the pan-gene families were categorized as genome-specific types (Figure 2c). For the C subgenome, a similar percentage of three kinds of pangene families was observed, although there was a slightly lower percentage of core gene families and a slightly higher percentage of dispensable gene families (Figure 2d).

Gene function enrichment analysis revealed that the core gene families were mainly involved in cytokinin-activated signaling pathway, glucose metabolism process, response to bacterium, mismatch repair, phosphorelay signal transduction system, and intracellular protein transport; while the indispensable gene families were associated with phosphate ion transport, pollen recognition, cell surface receptor signaling pathway, positive regulation of ubiquitin protein ligase activity, regulation of gene

expression, telomere maintenance and DNA integration (Figure S4), suggesting the different functions of the two classes of genes in rapeseed growth and development. However, there were no enriched GO terms with the same significance level for the species-specific gene families. Gene expression analysis on NTS57 revealed that the expression levels of core genes were significantly higher than those of dispensable genes in different tissues, and the specific genes had the lowest gene expression levels. (Figure S5, $P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test), reflecting the functional importance of the three groups of genes.

Extensive structural variations between NTS57 and other rapeseed varieties

The specific characteristics of NTS57 prompted us to further explore structural variations (SVs) at the genomic level with other rapeseed genomes. For this, purpose, we performed pairwise genome alignments between each of the nine rapeseed genomes and the NTS57 genome. We identified a total of 51,511 insertions, 58,775 deletions, 29,645 translocations, 15,654 duplications, and 4,047 inversions in the nine rapeseed accessions using NTS57 as the reference genome (Figure 3a). Among these different genomes, the artificially synthesized rapeseed line No2127 had the most SVs with the NTS57 reference genome, while the Tapidor had the least SVs. The total size of the SVs varied between 113.5 Mb and 153.6 Mb in the different accessions compared, with the length per SV ranging from a few tens of bp to hundreds of kb or even above the Mb scale (Figure S6). We found highly consistent patterns when using an alternative reference genome, ZS11, as the reference (Figure S7 and Figure S8). We also observed that insertions and deletions tended to be enriched at both chromosome ends and depleted near centromeres, with a distribution similar to gene density and opposite to GC content. In contrast to deletions and insertions, inversions, duplications and translocations appeared to be randomly distributed across the 19 chromosomes, although some enrichment was observed near centromeres (Figure 3b and Figure S9). These observations were also confirmed by calculating correlations between the chromosomal distributions of different SVs (Figure 3b).

Numerous studies have shown that SVs play an important role in plant growth and development (Alonge et al., 2020; Cai et al., 2021; Yuan et al., 2021). To reveal the potential functional influence of SVs in rapeseed, we mapped these SVs to NTS57 reference genes based on chromosomal positions and found that 39.2.6% (39,820) of protein-coding genes overlapped at least one SV across different species. GO analysis of these genes within SV hotspot regions revealed enrichment for peptidase activity, cysteine-type endopeptidase activity, and DNA integration, suggesting an important role of SVs in adaptation of the investigated rapeseed to abiotic stresses.

In addition to the SVs identified at the level of genomic fragments, the potential SVs based on the genome collinearity analysis at the level of protein-coding genes were also investigated. Overall, there are strong collinearity relationships between NTS57 and other rapeseed varieties, which do not reflect a large variation in chromosomal structure among different rapeseeds. However, we found a chromosomal fragment translocation from chromosome A03 to C03 in the NTS57 genome (Fig 3C). This translocation could not be a mis-assembly and this can be confirmed by chromosomal interaction and long read sequencing data analysis (Figure S10). Further inspection revealed that there are 1,283 protein-coding genes are located in this region. Among these genes, many are involved in plant temperature sensing or vernalization, including *VERNALIZATION INSENSITIVE 3* (*VIN3*) (Zhao et al., 2020), *Temperature-induced lipocalin-1*(*TIL-1*) (Huang et al., 2021), *Heat shock 70 kDa protein 10* (*HSP70-10*), *HSP90-2* (Jahan et al., 2023), and *Protein HEAT INTOLERANT 4* (*HIT4*) (Probst and Mittelsten Scheid, 2015), *etc*. (Table S5).

Gene expression of cold-stress regulated genes in NTS57

To identify the genes associated with cold tolerance in NTS57, we performed transcriptomic data analysis under normal and cold stress conditions that had been generated in a previous study (Wei et al., 2021). For this, the leaf RNA-Seq data from normal condition (CK), 12 h (T1), and 24 h (T2) after 4°C were mapped onto the NTS57 reference genome and gene expression levels of all protein-coding genes were calculated. Compared with the normal condition, we identified 15,888 and 22,479

differentially expressed genes (DEGs) in T1 and T2, respectively (Figure 4a, Table S6 and S7). Obviously, the number of identified cold-responsive genes increased with increasing duration of cold stress. These DEGs were highly reproducible by using different statistical analysis programs (Figure S11). Among these DEGs, there are 12,346 common DEGs at both time points after cold stress (Figure 4b), suggesting that these genes may represent core molecular components in response to cold stress.

Next, to test whether the DEGs are closely related to cold response, we checked their possible gene functions through gene function annotation and literature mining. Indeed, we observed that many of them are directly involved in plant cold response or cold acclimation process, including genes encoding superoxide dismutase (SOD) (Lin et al., 2019), VERNALIZATION INSENSITIVE 3(VIN3) (Zhao et al., 2020), transcription factor LUX (Chow et al., 2014), RNA-binding protein CP29B (Kupsch et al., 2012), flavanone 3-dioxygenase 3 (Wu et al., 2023), and dehydration-responsive element-binding protein 2E (DREB2E) (Agarwal et al., 2017), etc. Also, many genes annotated as cold response or cold induced genes are observed among these DEGs (Table S6 and S7). For example, there were 84 genes directly annotated as cold response or cold induced genes in the NTS57 reference genome, of which 38 and 49 were identified as DEGs in T1 and T2, respectively, implying that these DEGs were enriched in the high-confidence reference gene set for rapeseed cold resistance research. In addition, by combining the pan-genome analysis, we identified 83 and 82 DEGs in T1 and T2 samples, respectively, which are only present in the NTS57 genome, representing NTS57-specific genes functioning in cold response (Table S8).

To further understand the possible biological functions of the DEGs, we performed Gene Ontology and KEGG pathway enrichment analysis. It revealed that at the biological process (BP) level of DEGs in T1 are involved in the regulation of circadian rhythms, photosynthesis, cold response and xyloglucan metabolism (Figure 4c). At the cellular component (CC) level, these DEGs are associated with MCM complex, photosystem and cell wall, etc. (Figure S12). At the molecular function (MF) level, these DEGs play roles in UDP-glycosyltransferase, protein serine/threonine phosphatase activity, and ADP binding, etc. (Figure S12). For DEGs in T2 sample, we

observed a similar enrichment result, suggesting that cold-responsive genes are relatively stable with increasing cold treatment time (Figure S12). The KEGG pathway enrichment results showed that the DEGs were mainly involved in MAPK signaling pathway, circadian rhythm, flavonoid biosynthesis, photosynthesis, etc., which were similar to the results of GO enrichment analysis (Figure S13).

To further find the possible functional modules controlling cold resistance in NTS57, a weighted gene co-expression network analysis (WGCNA) was performed by integrating the above cold stress RNA-Seq data and a newly generated gene expression atlas of NTS57 from different tissues and developmental stages (Figure S14, S15 and Table S9). Based on the WGCNA, a total of 25 distinct co-expression modules with 24 to 15,396 genes per module were identified, and 15,496 uncorrelated genes were assigned to a grey module that was ignored in the following study (Figure 4d). We then mapped the 12,346 union DEGs onto these modules and tested whether these DEGs are overrepresented in some particular modules. Indeed, there are 10 modules, except the gray module, that are enriched in DEGs (hypergeometric test; $P < 0.05$, Figure 4e), suggesting that the DEGs function cooperatively with the gene in these modules. To confirm that these 10 modules are associated with cold stress, we performed a module–trait relationships analysis and found that the "darkgray" (r = 0.55, P = 3e-05), "darkgreen" (r =0.45, P = 0.001), "lightcyan" (r = 0.59, P = 6e-6), and "brown" (r= -0.67, P = 8e-08) modules were highly correlated with the cold stress (Figure 4f). Thus, we focus on the common modules, including "darkgreen", "lightcyan", and "brown", which were both enriched in DEGs and correlated with cold stress. The high correlation of these three modules with cold stress was also confirmed by gene expression heat map analysis (Figure S16-18). Gene function enrichment analysis revealed that genes in these modules are involved in protein folding and heat response, fatty acid biosynthetic process, guard cell differentiation, transcription, DNA-templated, and regulation of transcription by RNA polymerase II or epigenetic modification, etc. (Figure S19). Collectively, the diverse biological functions of these modules suggest that the cold response of rapeseed is involved in different biological processes beyond our previous knowledge.

DNA hypomethylation contributes to cold resistance in the NTS57

Most of the existing research has documented that cold stress induces genomic demethylation in many plant species, including rape, maize, wheat, Arabidopsis, and *Antirrhinum majus* (Peng and Zhang, 2009). However, it is unclear whether this rule applies to the materials of interest in this study. Therefore, we calculated genome-wide DNA methylation levels around protein-coding genes in the NTS57 genome under cold stress and normal conditions. The results showed a significant decrease in methylation levels across the gene body and flanking regions after cold stress compared to normal conditions in all cytosine sequence contexts CpG, CHG and CHH (H for A, T or C) (Figure 5a). In particular, the decrease in CHH methylation was most pronounced, which was followed by CHG and CpG methylation, suggesting that CHH methylation is more sensitive to cold stress. Furthermore, when the A and C subgenomes were treated independently, the similar trends were observed, although the C subgenome had an overall higher methylation level compared to A, which was independent of the treatment conditions of the samples (Figure S20), which is consistent with previous studies (Bird et al., 2021; Chalhoub et al., 2014).

In plants, DNA methylation is enriched over heterochromatic transposable elements (TEs) and repetitive elements (Gallego-Bartolome, 2020). We next analyzed DNA methylation changes in repeat sequences and transposons in the NTS57 genome. First, as expected, there was a high enrichment of DNA methylation on repetitive sequences including LTR and DNA transposons and tandem repeats in general, although we noticed two exceptions that CHG and CHH methylation in tandem repeats showed some deletions compared to the surrounding regions (Fig 5b). We also found that methylation levels were decreased in all repeats and in all methylation contexts under cold stress, which was very similar to that of the genic regions, suggesting that methylase acts on genes or repeats in a similar way under plant cold stress response.

To further understand the methylation dynamics upon cold stress, we identified differentially methylated regions (DMRs) between the two cold treatment samples (T1 and T2) and the CK. A total of 631-9,004 DMRs were identified in all sequence contexts upon cold stress (Figure 5c). Obviously, the number of hypo-methylation regions is

higher than that of the hyper-methylation regions in both T1 and T2 samples, which is consistent with the observation from methylation profiles that the methylation level decreased after cold treatment (Figure 5a). In addition, we observed that 47%-61% of these DMRs were located in the distal intergenic regions, followed by promoter regions of genes ranging from 24.12%-35.47% (Figure 5d), suggesting that the non-coding regions are the potential methylation target sites that response to cold stress in rapeseed. In particular, the DMRs in the CpG context were more likely to be deposited in promoter regions than those in the CHH and CHG in turn. On the contrary, DMRs in the CHG context were more frequently located in the distal intergenic region that in the CHH and CpG contexts in turn. Based on these DMRs, we obtained 24,947 associated protein-coding genes, hereafter referred to as DMR-associated genes (Figure S21). Consistent with the DMRs number, the percentage of hypomethylated genes (63.5%-72.5%) was much higher than the number of hypermethylated genes in all three methylation contexts in the two samples of cold treatment. After comparing these DMR-linked genes with DEGs from two cold treatment samples, there are 506~2,381 common genes in three sequence contexts (Figure S22). We then determined the fraction of genes showing differential (hypo- or hyper-) methylation in different sequence contexts and their direction of differential expression (up or down-regulated) upon cold stress. In general, there was no consistent relationship between differential methylation in different sequence contexts and the direction of differential gene expression (Figure S23), suggesting that there is a complex DNA methylation regulatory mechanism regarding to DEGs at low temperature stress in rapeseed.

GWAS analysis identifying cold-stress related gene loci in rapeseed

Previous studies have shown that the REC increases due to the increase in cytoplasmic membrane permeability and extravasation of soluble substances when plants are exposed to cold stress (Lu et al., 2017). Therefore, the level of REC can be used as an index to evaluate the cold resistance of plants (Yu et al., 2021). To identify excellent loci controlling cold resistance in rapeseed, 289 individuals with different eco-types were obtained from a previous study (Xuan et al., 2020). These individuals were

originally derived from 24 countries (Figure 6a and Figure S24), and represent the core rapeseed germplasm, covering 97.3% of the SNPs and 97.9% of the InDels of the 991accession collection (Wu et al., 2019; Xuan et al., 2020). To measure the cold tolerance of these germplasm, REC was measured with three biological replicates. Statistical analysis of the phenotype data showed that the data followed a normal distribution (Figure 6b and c). After mapping the re-sequencing data to the NTS57 reference genome, we identified a total of 9,446,761 high-quality single nucleotide polymorphisms (SNPs) and 469,176 insertions and deletions (InDels) with minor allele frequencies (MAFs) >0.05 (Figure S25 and Figure S26). The ratio of nonsynonymous to synonymous substitutions was 0.69, which was comparable to that of a previous study using 418 diverse rapeseed accessions (Hu et al., 2022). The PCA plot of the first two components (PC1 and PC2), based either on SNP or InDels, could separate the 290 accessions into three groups that roughly corresponding to three ecotypes of rapeseed (Figures 27).

The GWAS analysis identified 26 SNPs that located on nine chromosomes that are strongly associated with REC. Among these SNPs, 15 were tightly localized on chromosome C03 and C07, and the remaining 11 SNPs were scattered on other seven chromosomes (Figure 6d and Table S10), suggesting that the two major genomic regions are associated with cold resistance in rapeseed evolution or breeding. We also obtained six loci associated with REC by InDel-GWAS analysis, five of which overlapped with the results of the SNP-GWAS analysis (Figure S28 and Table S11). The high overlap rate based on the two polymorphisms analyses reflects the reliability of the identified loci controlling rapeseed cold resistance. It should also be noted that the GWAS loci in chromosome C07 were missed by InDel-GWAS, probably due to the lower number of InDels in the region. The effect of each locus on the REC under cold stress was evaluated to further confirm the reliability of these candidate loci (Figure S29 and S30).

Next, we focused on these loci and explored the associated genes based on the SNP and InDels annotation information. For the chromosome C03 loci containing 10 closely linked SNPs, there were three associated protein-coding genes, including *C03G039940*,

C03G039950, and *C03G039960* (Figure 6e). *C03G039940* encodes a NADH: ubiquinone oxidoreductase subunit A6 (NDUFA6) orthologous to AT3G12260 (NDUFA6 or B14) in *Arabidopsis.* Studies have shown that NDUFA6 belongs to the LYR protein superfamily and anchors an acyl carrier protein to complex I of the respiratory chain, forming a structurally well-defined subdomain essential for complex I activity (Angerer, 2013). *C03G039950* encodes a putative protein arginine N-methyltransferase (PRMT3) that is required for ribosome biogenesis by affecting pre-rRNA processing (Hang et al., 2021). *C03G039960* encodes a retinoblastoma homolog RETINOBLASTOMA-RELATED protein (RBR or RBR1). RBR controls nuclear proliferation in the female gametophyte and is also required for proper differentiation of male gametophytic cell types (Ebel et al., 2004).

For another ~60 kb significant loci in chromosome C07 identified by SNP-GWAS, there were 10 nine associated genes, including *C07G040420-C07G040510* arranged in this region (Figure 6f). Importantly, there are three genes, *C07G040480, C07G040490*, and *C07G040510*, that were deferentially expressed under cold stress based on our RNA-Seq analysis (Figure S31). *C07G040480* encodes an unnamed protein product which was annotated from *Brassica oleracea. C07G040490* encodes a transcription factor orthologous to Arabidopsis MYB34. *C07G040510* encodes a G-type lectin S-receptor-like serine/threonine-protein kinase (RLK1), which plays critical and diverse roles in plant growth, development, and stress response (Ye et al., 2017). For both MYB34 and RLK1, their expression levels were significantly induced under cold stress by qRT-PCR analysis (Figure S32). Overall, these candidate gene loci will be important genomic target of genetic modification to develop cold-resistant rapeseed cultivars.

Discussion

Cold tolerance is one of the most important traits in crop domestication and breeding. In this study, we generated an improved high-quality genome assembly of a prominent cold-resistant rapeseed cultivar, NTS57. Various genome assembly indexes were used to confirm the high genome completeness. For example, the contig N50 of the present assembly is 21.81 Mb, almost 10 times longer than that of previous studies (Song et al.,

433 2020), with the BUSCO score as high as 99.19%. Moreover, the number of assembly
434 gaps is remarkably reduced in the present reference genome. Thus, this chromosome-
435 level reference genome can allow us to systemically analyze the cold tolerance of
436 rapeseed.

437 At the phylogeny level, the A and C subgenomes of the NTS57 have different
438 clustering patterns, where the A subgenome was more likely to form a clade with that
439 of semi-winter rapeseed, but the C subgenome was more closely related to the winter
440 type rapeseed. This incongruity could be explained by the unique breeding history of
441 the cold-tolerant NTS57. Basically, NTS57 underwent an additional artificial
442 interspecific hybridization between winter rapeseed and *Brassica rapa* after being bred
443 from winter type rapeseed. This interspecific hybridization led to a high sequence
444 similarity between the A subgenome of NTS57 and that of semi-winter rapeseed.
445 However, the C subgenome showed sequence conservation with that of the winter
446 rapeseed. At the pan-genome level, there are extensive genomic SVs between NTS57
447 and other *rapeseed v*arieties, which may contribute to the local environmental
448 adaptations.

449 Studies have shown that Arabidopsis can respond to cold stress through either
450 DREB1/CBF or DREB1/CBF-independent signaling pathways (Kidokoro et al., 2022).
451 In the classical DREB1/CBF pathway, DREB1/CBF were considered as key
452 transcription factors that involved in cold-responsive gene expression (Shi et al., 2018).
453 By analyzing the transcriptomic data, there are at least 16 *DREBs* homologous genes in
454 NTS57 that are induced by cold stress (Table S6 and S7). ICE1 has been considered as
455 one of many transcription factors that can regulate the *DREB1* expression (Guo et al.,
456 2018). However, we did not observe a significant change in the expression of *ICE1* in
457 the NTS57 after cold stress, suggesting that ICE1 may not regulate the expression of
458 the *DREB1* genes, or at least ICE1 is not induced by low temperature. Therefore, the
459 ICE1-DREB1 regulatory module needs to be further confirmed (Kidokoro et al., 2022).
460 In addition to the DREB1/CBF pathway, other transcription factors such as RVEs,
461 ZAT12, HSFA1, BZR1, etc. are also differentially expressed under cold stress in the
462 NTS57, suggesting a lot of crosstalk between DREB1/CBF and the DREB1/CBF-

independent signaling pathways. Also, several uncharacterized genes such as *A07G034200*, *C08G048350*, *C01G004880* showed extremely significant expression changes upon cold stress. Whether they belong to DREB1/CBF or DREB1/CBF-independent signaling pathways needs further research, and these genes will be new potential targets for genetic modification of rapeseed to improve cold tolerance.

As one of the largest TF families, MYB TFs play important roles in biotic and abiotic stress responses in plants, including cold stress responses (Chinnusamy et al., 2007; Dubos et al., 2010). For example, MYB14 and MYB15 negatively regulate cold tolerance by repressing the expression of CBF genes in Arabidopsis (Agarwal et al., 2006; Chen et al., 2013). In rice, however, overexpression of MYB4, MYB3R-2, MYBS3, and MYB2 could enhance plant cold tolerance (Dai et al., 2007; Su et al., 2010; Vannini et al., 2004; Yang et al., 2012). Recent studies in apple showed that some MYB members are positive regulators of cold tolerance (An et al., 2020; Jiang et al., 2022; Xie et al., 2018), while some negatively regulate cold tolerance in plants (Wu et al., 2018; Xu et al., 2018a; Xu et al., 2018b). Using GWAS analysis in this study, we identified a homologous gene encoding MYB34. The genetic variation located at this locus was significantly associated with cold tolerance in rapeseed population, which is important for improving the cold tolerance of rapeseed.

In summary, we have dissected the genomic architecture of cold tolerance by generating a high-quality reference genome of the NTS57 and performing phylogenetic, pan-genomic, DNA methylomic, and transcriptomic analyses of cold-responsive genes. These resources and findings provide valuable information that can be used in the breeding of rapeseed with high cold tolerance.

Materials and Methods

Plant materials and DNA extraction for genome assembly

To perform genome sequencing and assembly of rapeseed NTS57, the seeds were successfully germinated, and plants were grown in the greenhouse of Gansu Agricultural University under the growth condition of 20°C/18°C and 16/8 h days. Leaf tissue was collected from 4-week-old plants, flash-frozen and stored at -70°C. High

molecular weight genomic DNA was extracted using the CTAB method, followed by purification using the Grandomics Genomic Kit for regular sequencing according to the manufacturer's standard operating procedure. The DNA degradation and contamination of the extracted DNA was checked on 0.75% agarose gels. DNA purity was then detected using NanoDrop™ One UV-Vis spectrophotometer (Thermo Fisher Scientific, USA), of which OD260/280 was between 1.8 and 2.0 and OD 260/230 is between 2.0-2.2. Finally, the DNA concentration was further measured using Qubit® 3.0 Fluorometer (Invitrogen, USA).

PacBio HiFi library preparation and DNA sequencing

For long-read sequencing, the extracted DNA was sequenced on a PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA) using the circular consensus sequencing (CCS) model from two SMRT cells using 15kb preparation solutions. The main steps for library preparation are: gDNA shearing, DNA damage repair, end repair and A-tailing, ligation with hairpin adapters from the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences), nuclease treatment of the SMRTbell library with the SMRTbell Enzyme Cleanup Kit, size selection, and binding to polymerase.

Briefly, a total amount of 8 μg DNA per sample was used for the DNA library preparation. The genomic DNA sample was first sheared with g-TUBEs (Covaris, USA) according to the expected size of the fragments for the library. Single strand overhangs were removed, and DNA fragments were damage repaired, end repaired, A-tailed, and ligated with the hairpin adaptor for PacBio sequencing. The library was treated by nuclease with SMRTbell Enzyme Cleanup Kit and purified with AMPure PB Beads. Target fragments were screened using PippinHT (Sage Science, USA). The SMRTbell library was purified with AMPure PB Beads and Agilent 2100 Bioanalyzer (Agilent Technologies, USA) was used to determine the size of the library fragments. Finally, sequencing was performed on a PacBio Sequel II instrument using Sequencing Primer V5 and Sequel II Binding Kit 2.2 at the Genome Center of Grandomics (Wuhan, China). After sequencing, the resulting BAM files were submitted to the pbccs program (https://ccs.how/) with default parameters, resulting in 65.16Gb HiFi pass reads with

an average length of 15.6 kb.

Preparation of short reads libraries and DNA sequencing

Short-read libraries were prepared using the MGIEasy Universal DNA Library Prep Kit V1.0 (CAT#1000005250, MGI) according to the standard protocol. Briefly, 1 μg of genomic DNA was randomly fragmented by Covaris. The fragmented DNA was selected to an average size of 200-400 bp using MGIEasy DNA Clean Beads (CAT#1000005279, MGI). The selected fragments were then end-repaired, 3' adenylated and then ligated with adapters. The DNA samples were amplified by PCR and the products were purified using MGIEasy DNA Clean Beads (CAT#1000005279, MGI). The double-stranded PCR products were heat denatured and circularized using the splint oligo sequence in the MGIEasy Circularization Module (CAT#1000005260, MGI). The single-strand circle DNA (ssCir DNA) was formatted as the final library and qualified by QC. The qualified libraries were sequenced on DNBSEQ-T7RS platform.

Transcriptome library preparation and sequencing

Different tissues at different developmental stages were sampled to perform RNA-Seq with three biological replicates (Supplementary Table S9). Briefly, total RNA was extracted for each sample by grinding the tissue in TRIzol reagent (TIANGEN) on dry ice and processed according to the manufacturer's protocol. Poly-A RNAs were enriched from total RNA using Dynabeads mRNA Purification Kit (Cat. No. 61006, Invitrogen) and fragmented into small pieces using fragmentation reagent in MGIEasy RNA Library Prep Kit V3.1 (Cat# 1000005276, MGI). First-strand cDNA was synthesized using random primers and reverse transcriptase, followed by second-strand cDNA synthesis. The synthesized cDNA was end-repaired, A-tailed and ligated to the sequencing adapters according to the library construction protocol. The cDNA fragments were amplified by PCR and purified using MGIEasy DNA Clean Beads (CAT#1000005279, MGI). The library was analyzed on the Agilent Technologies 2100 Bioanalyzer. The double-stranded PCR products were heat denatured and circularized using the splint oligo sequence in the MGIEasy Circularization Module (CAT#1000005260, MGI). The single-strand circle DNA (ssCir DNA) were formatted

as the final library. The qualified libraries were sequenced on DNBSEQ-T7RS platform.

Hi-C library preparation and sequencing

Freshly harvested leaves from the 4-week-old plants were cut into 2 cm pieces and vacuum infiltrated in nuclear isolation buffer supplemented with 2% formaldehyde. Cross-linking was stopped by the addition of glycine and additional vacuum infiltration. The fixed tissue was then ground to powder and resuspended in nuclear isolation buffer to obtain a nuclear suspension. The purified nuclei were digested with 100 units of DpnII and labeled by incubation with biotin-14-dCTP. Biotin-14-dCTP was removed from unligated DNA ends due to the exonuclease activity of T4 DNA polymerase. The ligated DNA was sheared into 300-600 bp fragments, and then was blunt-end repaired and A-tailed, followed by purification by biotin-streptavidin-mediated pull-down. Finally, the Hi-C libraries were quantified and sequenced using the MGI-2000 platform.

Genome size and heterozygosity estimation

To understand the genomic characteristics of the NTS57, the k- was removed was removed analysis was performed using Illumina DNA data prior to genome assembly to estimate the genome size and heterozygosity. Briefly, quality-filtered reads were subjected to 25-mer frequency distribution analysis using the KMC program (Kokot et al., 2017). By analyzing the 25-mer depth distribution from the clean reads using FindGSE (Sun et al., 2018) and GenomeScope2.0 (Ranallo-Benavidez et al., 2020), we estimated the genome size using the following equation: G = K-num/K-depth (where K-num is the total number of 25-mers, K-depth is the k-mer depth, and G is the genome size). Finally, the heterozygosity and repeat content of the NTS57 genome were estimated by combining the simulation data results of Arabidopsis with different heterozygosity and the frequency peak distribution of 25-mers.

Genome denovo assembly and completeness assessment

The high quality HiFi reads were first used to assemble the initial contigs using the Hifiasm (v.0.13-r308) package with default parameters (Cheng et al., 2021). To improve the accuracy of the assembly, the assembled contigs were polished with Nextpolish (Hu

578 et al., 2020) using short reads with default parameters. Finally, redundancies were
579 removed using Haplomerger2 (Huang et al., 2017) to generate the primary contigs.
580 Next, Hi-C reads were used to support the assembly of the genome at the
581 chromosome level. For this, the low-quality Hi-C reads (quality scores<20), adaptor
582 sequences, and sequences shorter than 30 bp were filtered out using fastp (Chen et al.,
583 2018), and then the clean paired-end reads were mapped to the assembled contig
584 sequences using bowtie2 (v2.3.2) (Langmead and Salzberg, 2012) (-end-to-end --very-
585 sensitive -L 30) to obtain the uniquely mapped paired-end reads. Valid interaction
586 paired reads were identified and rerained by HiC-Pro (v2.8.1) (Servant et al., 2015)
587 from the unique mapped paired-end reads for further analysis. Invalid read pairs,
588 including dangling-end, self-cycle, re-ligation, and dumped products were filtered.
589 Scaffolds were further clustered, ordered, and oriented on chromosomes using
590 LACHESIS (Burton et al., 2013) with parameters CLUSTER_MIN_RE_SITES=100,
591 CLUSTER_MAX_LINK_DENSITY=2.5, CLUSTER NONINFORMATIVE RATIO
592 = 1.4, ORDER MIN N RES IN TRUNK=60, ORDER MIN N RES IN SHREDS=60.
593 Finally, placement and orientation errors with obvious discrete chromatin interaction
594 patterns were manually adjusted.
595 The completeness of the genome assembly was evaluated using BUSCO v4.0.5
596 (Benchmarking Universal Single Copy Orthologs) and CEGMA (Core Eukaryotic
597 Gene Mapping Approach) (Parra et al., 2007; Simao et al., 2015). To assess the accuracy
598 of the assembly, all the Illumina paired-end reads were mapped to the assembled
599 genome using BWA software (Li and Durbin, 2009), and the mapping rate and genome
600 coverage of the sequencing reads were assessed using SAMtools v0.1.1855 (Li et al.,
601 2009). Furthermore, the base accuracy of the assembly was calculated using bcftools
602 (Danecek and McCarthy, 2017). In addition, the coverage of potentially expressed
603 genes of the assembly was examined by aligning RNA-Seq data to the assembly using
604 HISAT2 with default parameters (Kim et al., 2019). To avoid including mitochondrial
605 sequences in the assembly, the draft genome assembly was submitted to the NCBI (NT)
606 library and the aligned sequences were eliminated.

Repetitive DNA elements and gene model annotation

Tandem repeats were first annotated using the software GMATA (Wang and Wang, 2016) and Tandem Repeats Finder (TRF) (Benson, 1999), where GMATA identifies the simple sequence repeats (SSRs) and TRF recognizes all tandem repeat elements in the entire genome. Transposable elements (TEs) in the NTS57 genome were then identified using a combination of de novo and homology-based strategies. For de novo prediction, three de novo prediction programs, MITE-hunter (Han and Wessler, 2010), RepeatModeler (Flynn et al., 2020), and LTR_retriever (Ou and Jiang, 2018), were used with default parameters. The results from these programs were integrated into a de novo database, which was then aligned to Repbase (http://www.girinst.org/repbase) and classified into different types. Using the resulting database, repetitive sequences were identified by homologous sequence search using RepeatMasker with default parameters (Tarailo-Graovac and Chen, 2009). Overlapping repeat elements belonging to the same repeat class were merged as a final non-redundant repeat sequence annotation. For centromere sequences identification, the quarTeT program was used with the NTS57 genome sequences as the input with default parameters (Lin et al., 2023). The top candidates were selected based on the length and periods of repetitive sequences. Meanwhile, the gene density and LTR density distribution across chromosomes were also used to help to determine the centromere positions in chromosomes.

For gene model prediction, three independent approaches, including ab initio prediction, homology search, and reference-guided transcriptome assembly, were used to predict genes in the repeat-masked genome. Specially, GeMoMa (Keilwagen et al., 2019) was used to align the homologous peptides from related species to the assembly and then obtain the gene structure information, which was the homology prediction. For RNA-Seq-based gene prediction, filtered RNA-Seq reads were mapped to the reference genome using STAR with default parameters (Dobin et al., 2013). The transcripts were then assembled using stringtie (Pertea et al., 2015) and open reading frames (ORFs) were predicted using PASA (https://github.com/PASApipeline/PASApipeline). For the de novo prediction, RNA-Seq reads were de novo assembled using stringtie (Pertea et al., 2015) and analyzed

using PASA to generate a training set. Augustus with default parameters (Stanke et al., 2004) was then used for ab initio gene prediction with the training set. Finally, EVidenceModeler (EVM) (Haas et al., 2008) was used to generate an integrated gene set from which genes with TE were removed using the TransposonPSI package (http://transposonpsi.sourceforge.net/) and the miscoded genes were further filtered. Untranslated regions (UTRs) and alternative splicing regions were determined using PASA based on RNA-seq assemblies. We retained the longest transcripts for each locus, and regions outside of the ORFs were designated as UTRs.

Functional annotation of gene models

Functional information of genes, motifs and domains of their proteins were assigned by comparison with public databases including Swiss-Prot (Boutet et al., 2007), nonredundant protein database NR (Pruitt et al., 2007), KEGG (Kanehisa and Goto, 2000), KOG (Tatusov et al., 2001) and Gene Ontology (Gene Ontology, 2015). The putative domains and GO terms of genes were identified using the InterProScan program with default parameters (Quevillon et al., 2005). For the other four databases, BLASTp (Altschul et al., 1990) was used to compare the EvidenceModeler-integrated protein sequences against the four well-known public protein databases with an E-value cutoff of 1e-05, and the results with the hit with the lowest E-value were retained. The results from the above five database searches were concatenated.

Annotation of non-coding RNAs (ncRNAs)

Two strategies were used to obtain the ncRNA (non-coding RNA): database search and model prediction. Transfer RNAs (tRNAs) were predicted using tRNAscan-SE with eukaryote parameters (Chan et al., 2021). MicroRNA, rRNA, small nuclear RNA, and small nucleolar RNA were detected using Infernal cmscan to search the Rfam database (Nawrocki and Eddy, 2013). The rRNAs and their subunits were predicted using RNAmmer (Lagesen et al., 2007).

Synteny analysis between NTFS57 and other rapeseeds

MCscan (Python version) (Tang et al., 2008) was used to identify collinear blocks of at least of 30 genes in NTS57 and other rapeseed genomes.

Phylogenetic analysis

To determine the phylogenetic relationship between NTS57 and other published rapeseed cultivars, the phylogenetic tree was reconstructed using 10 rapeseed species and two ancestral reference genomes, *B. rapa* Chiifu (v4.0) and *B. oleracea*, with *A. thaliana* as the outgroup species. For this, the single-copy orthogroups of the above 13 species were firstly identified using amino acid sequences by OrthoFinder v2.0.9 with the parameter "-M msa" (Emms and Kelly, 2019). Then, the multiple sequence alignment in each orthogroup was performed using mafft v7.453 (Katoh and Standley, 2013), and the resulting alignments were concatenated into a superalignment. Finally, the phylogenetic tree was determined by the maximum likelihood (ML) method as implemented in IQ-TREE v1.6.9 with 1000 bootstrap tests (Nguyen et al., 2015), and the best-fit model was determined as being JTT+F+R4 by ModelFinder with free rate heterogeneity (Kalyaanamoorthy et al., 2017).

Gene expression analysis

For both the newly generated RNA-seq datasets and published RNA-seq datasets, Trimmomatic (v0.36) (Bolger et al., 2014) with default parameters was used to remove adapters and low-quality reads from the raw RNA-seq reads. The clean RNA-seq reads were mapped to the corresponding reference genome using Hisat2 with default settings (Kim et al., 2019). Gene expression abundance was estimated using FPKM (fragments per kilobase of exon per million fragments calculated by StringTie (v2.1.6) (Pertea et al., 2015) for each transcript. The gene co-expression network was constructed using the R package WGCNA (Langfelder and Horvath, 2008) by using 51 RNA-Seq samples from different tissues and developmental stages as well as cold stress treatment.

RT-qPCR experiments

Total RNA was extracted from the developing leaves of NTS57 under normal conditions (CK) or 12 h after -4°C treatment using the TRIzol method (Tiangen, Beijing, China) according to the manufacturer's instructions. The gene *BnACTIN* was used as

the internal control. The primers used in the RT-qPCR are listed in Table S12. The relative quantification ($2^{-\Delta\Delta Ct}$) of gene expression was evaluated using comparative cycle threshold method, and each sample was replicated for three times.

Analysis of whole-genome bisulfite sequencing data

Whole genome methylation data for leaf tissues of NTS57 under normal or low temperature (-4°C) were reanalyzed from our previous study (NCBI accession: PRJNA685002) (Wei et al., 2022). Briefly, fastp was first used to perform quality filtering, FastQC v.0.11.7 (Wingett and Andrews, 2018) was used for quality control and Bismark (v.0.23.0) (Krueger and Andrews, 2011) was used to align the trimmed reads to our assembled reference genome of NTS57. The DNA methylation levels for individual CpGs, CHG and CHH were calculated separately using methyKit (v.1.22.0) (Akalin et al., 2012), and we retained only cytosine sites with ≥10 covered sequencing reads. Methylated reads (containing Cs) and unmethylated reads (containing Ts) at each cytosine site were counted, and the percentage of methylated reads among the total reads covering the corresponding cytosine was calculated to quantify the DNA methylome for each sample at the single base resolution. To identify the DMRs, we used a window-based method to calculate the overall methylation level between samples for comparisons. For this purpose, 1,000 bp continuous windows were generated across 19 chromosomes of the NTS57 genome, and the windows covering no less than 10 reads that showed ≥25% methylation level difference with q-value ≤0.01 were considered as DMRs using methyKit (v.1.22.0) (Akalin et al., 2012).

To compare the overall methylation level for different genomic features, such as upstream, downstream, the gene body of protein-coding genes, and repeat elements, the mean CpG, CHG, and CHH methylation levels for 2 kb upstream and downstream of each type of the genomic features were binned into 100 equal-length windows, and the mean methylation levels in each window were calculated using a custom R script.

SV identification

For SV calling, minimap2 (v.2.21-r1071) (Li, 2018) was used to generate pairwise genome alignments between each of the nine rapeseed accessions and the NTS57

genome using the parameters "-ax asm5 -eqx". The alignment results were then passed to SyRI (Goel et al., 2019) and SVs consisting of insertions, deletions, inversions (<1 Mb in size), translocations (>50 bp in length) and duplications were retained. SVs containing 'N' sequences and the ambitious alignment margins and/or poor synteny alignment around the breakpoint were filtered. SVs between different genomes were visualized using plotsr (Goel and Schneeberger, 2022).

Phenotyping of GWAS population

289 core accessions of rapeseed germplasm were selected from a global germplasm collection of 991 accessions based on phylogenetic and principal component analyses (Xuan et al., 2020). The 289 core accessions were grown in Yangling, Shaanxi, China (34°27'N and 108°06'E) during the overwintering from September 2021 to January 2022. The genetic material was sown in nursery beds in plastic tunnels and then transplanted to the open field. For each accession, the REC was measured on the fourth fully expanded leaf of five mixed plants.

SNP calling and GWAS

Raw sequencing reads were obtained from (Wu et al., 2019), and trimmed using Trimmomatic (v0.36) (Bolger et al., 2014). The clean reads were then aligned to the NTS57 reference genome using bwa-mem (Li and Durbin, 2009). The alignment results were sorted, and duplicate reads were removed using (Tarasov et al., 2015). SNPs and InDels were called at the population level using the HaplotypeCaller module in GATK v4.2.2.0 (McKenna et al., 2010), after which they were filtered using the parameters: QD < 2 || QUAL < 30 || FS > 60 || MQ < 40 || MQrankSum < -12.5 || ReadPosRankSum < -8.0. Finally, only SNPs or InDels with a MAF greater than 5% and missing data less than 50% were considered as high-quality SNPs or InDels. The Association analysis was performed using GEMMA v0.98.5 (Zhou and Stephens, 2012). The *P* value of each SNP was calculated, and -log10(*P*) > 6 was defined as the suggestive threshold.

Acknowledgements

This work was supported by the Natural Science Foundation of Gansu Province (22JR5RA862), the Gansu Province Higher Education Institutions Young Doctoral

Fund (2023QB-126), the Scientific Research Staring Foundation of Gansu Agricultural University (GAU-KYQD-2020-27), the Gansu Provincial Science and Technology Major Project (22ZD6NA009), the Central-Guided Local Science and Technology Development Fund (ZCYD-2021-17), and the National Natural Science Foundation of China (32360520). We thank Lixin Jang from Zhejiang University, Hangzhou, for providing the rapeseed seeds used in the GWAS analysis. We also thank Tao Yan from Hunan Agricultural University, Changsha, for critically reading of this manuscript.

Author contributions

Z.L. and Z.W. designed the project; G.Z., S.Y., and X.D. prepared the materials; Z.W., G.Z., S.Y., X.D., H.L., and Z.H. performed the experiments; Z.W. performed the data analysis; Z.W. and Z.L. wrote the manuscript; Y.S., Y.W., Y.N., Z.H., and J.H. discussed and revised the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of interest

No conflict of interest was declared.

Data availability

All the raw sequencing data and the genome assembly generated for this research are archived at NCBI under accession number PRJNA1012021. Genome annotation files and other related documents are available in figshare (https://doi.org/10.6084/m9.figshare.24098538.v1) or the database (http://oilseed.online:8000/). All the materials and related analysis scripts from this study are available upon request.

References

Agarwal, M., Hao Y Fau - Kapoor, A., Kapoor A Fau - Dong, C.-H., Dong Ch Fau - Fujii, H., Fujii H Fau - Zheng, X., Zheng X Fau - Zhu, J.-K. and Zhu, J.K. (2006) A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J Biol Chem* **281**, 37636-37645.

Agarwal, P.K., Gupta, K., Lopato, S. and Agarwal, P. (2017) Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance. *J Exp Bot* **68**, 2135-2148.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A. and Mason, C.E.

782 (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA
783 methylation profiles. *Genome Biol* **13**, R87.
784 Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H.*, et al.* (2020) Major Impacts
785 of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell*
786 **182**, 145-161 e123.
787 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool.
788 *J Mol Biol* **215**, 403-410.
789 An, J.P., Wang, X.F., Zhang, X.W., Xu, H.F., Bi, S.Q., You, C.X. and Hao, Y.J. (2020) An apple MYB
790 transcription factor regulates cold tolerance and anthocyanin accumulation and undergoes
791 MIEL1-mediated degradation. *Plant Biotechnol J* **18**, 337-353.
792 Angerer, H. (2013) The superfamily of mitochondrial Complex1_LYR motif-containing (LYRM) proteins.
793 *Biochem Soc Trans* **41**, 1335-1341.
794 Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**,
795 573-580.
796 Bird, K.A., Niederhuth, C.E., Ou, S., Gehan, M., Pires, J.C., Xiong, Z., VanBuren, R.*, et al.* (2021) Replaying
797 the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*.
798 *New Phytol* **230**, 354-371.
799 Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence
800 data. *Bioinformatics* **30**, 2114-2120.
801 Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. and Bairoch, A. (2007) UniProtKB/Swiss-Prot.
802 *Methods Mol Biol* **406**, 89-112.
803 Browse, J. and Xin, Z. (2001) Temperature sensing and cold acclimation. *Curr Opin Plant Biol* **4**, 241-246.
804 Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O. and Shendure, J. (2013) Chromosome-
805 scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat*
806 *Biotechnol* **31**, 1119-1125.
807 Cai, X., Chang, L., Zhang, T., Chen, H., Zhang, L., Lin, R., Liang, J.*, et al.* (2021) Impacts of
808 allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*.
809 *Genome Biol* **22**, 166.
810 Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J.*, et al.* (2014) Early
811 allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950-
812 953.
813 Chan, P.P., Lin, B.Y., Mak, A.J. and Lowe, T.M. (2021) tRNAscan-SE 2.0: improved detection and functional
814 classification of transfer RNA genes. *Nucleic Acids Res* **49**, 9077-9096.
815 Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor.
816 *Bioinformatics* **34**, i884-i890.
817 Chen, X., Tong, C., Zhang, X., Song, A., Hu, M., Dong, W., Chen, F.*, et al.* (2021) A high-quality *Brassica*
818 *napus* genome reveals expansion of transposable elements, subgenome evolution and disease
819 resistance. *Plant Biotechnol J* **19**, 615-630.
820 Chen, Y., Chen, Z.L., Kang, J.Q., Kang, D.M., Gu, H.Y. and Qin, G.J. (2013) Regulates Cold Tolerance in.
821 *Plant Mol Biol Rep* **31**, 87-97.
822 Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. and Li, H. (2021) Haplotype-resolved de novo assembly
823 using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170-175.
824 Chinnusamy, V., Zhu, J. and Zhu, J.K. (2007) Cold stress regulation of gene expression in plants. *Trends*
825 *Plant Sci* **12**, 444-451.

826 Chow, B.Y., Sanchez, S.E., Breton, G., Pruneda-Paz, J.L., Krogan, N.T. and Kay, S.A. (2014) Transcriptional
827 regulation of LUX by CBF1 mediates cold input to the circadian clock in Arabidopsis. *Curr Biol*
828 **24**, 1518-1524.
829 Cui, P., Li, Y., Cui, C., Huo, Y., Lu, G. and Yang, H. (2020) Proteomic and metabolic profile analysis of low-
830 temperature storage responses in Ipomoea batata Lam. tuberous roots. *BMC Plant Biol* **20**, 435.
831 Dai, X., Xu, Y., Ma, Q., Xu, W., Wang, T., Xue, Y. and Chong, K. (2007) Overexpression of an R1R2R3 MYB
832 gene, *OsMYB3R-2*, increases tolerance to freezing, drought, and salt stress in transgenic
833 Arabidopsis. *Plant Physiol* **143**, 1739-1751.
834 Danecek, P. and McCarthy, S.A. (2017) BCFtools/csq: haplotype-aware variant consequences.
835 *Bioinformatics* **33**, 2037-2039.
836 Ding, Y., Shi, Y. and Yang, S. (2019) Advances and challenges in uncovering cold tolerance regulatory
837 mechanisms in plants. *New Phytol* **222**, 1690-1704.
838 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P*., et al.* (2013) STAR: ultrafast
839 universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.
840 Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L. (2010) MYB transcription
841 factors in Arabidopsis. *Trends Plant Sci* **15**, 573-581.
842 Ebel, C., Mariconti, L. and Gruissem, W. (2004) Plant retinoblastoma homologues control nuclear
843 proliferation in the female gametophyte. *Nature* **429**, 776-780.
844 Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative
845 genomics. *Genome Biol* **20**, 238.
846 Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C. and Smit, A.F. (2020)
847 RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl*
848 *Acad Sci U S A* **117**, 9451-9457.
849 Gallego-Bartolome, J. (2020) DNA methylation in plants: mechanisms and tools for targeted
850 manipulation. *New Phytol* **227**, 38-44.
851 Gene Ontology, C. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-1056.
852 Goel, M. and Schneeberger, K. (2022) plotsr: visualizing structural similarities and rearrangements
853 between multiple genomes. *Bioinformatics* **38**, 2922-2926.
854 Goel, M., Sun, H., Jiao, W.B. and Schneeberger, K. (2019) SyRI: finding genomic rearrangements and local
855 sequence differences from whole-genome assemblies. *Genome Biol* **20**, 277.
856 Guan, S., Xu, Q., Ma, D., Zhang, W., Xu, Z., Zhao, M. and Guo, Z. (2019) Transcriptomics profiling in
857 response to cold stress in cultivated rice and weedy rice. *Gene* **685**, 96-105.
858 Guo, X., Liu, D. and Chong, K. (2018) Cold signaling in plants: Insights into mechanisms and regulation. *J*
859 *Integr Plant Biol* **60**, 745-756.
860 Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O*., et al.* (2008) Automated
861 eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble
862 Spliced Alignments. *Genome Biol* **9**, R7.
863 Han, Y. and Wessler, S.R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat
864 transposable elements from genomic sequences. *Nucleic Acids Res* **38**, e199.
865 Hang, R., Wang, Z., Yang, C., Luo, L., Mo, B., Chen, X., Sun, J*., et al.* (2021) Protein arginine
866 methyltransferase 3 fine-tunes the assembly/disassembly of pre-ribosomes to repress
867 nucleolar stress by interacting with RPS2B in arabidopsis. *Mol Plant* **14**, 223-236.
868 Hu, J., Chen, B., Zhao, J., Zhang, F., Xie, T., Xu, K., Gao, G*., et al.* (2022) Genomic selection and genetic
869 architecture of agronomic traits during modern rapeseed breeding. *Nat Genet* **54**, 694-704.

870 Hu, J., Fan, J., Sun, Z. and Liu, S. (2020) NextPolish: a fast and efficient genome polishing tool for long-
871 read assembly. *Bioinformatics* **36**, 2253-2255.
872 Huang, Q.X., Liao, X.Q., Yang, X.H., Luo, Y.C., Lin, P., Zeng, Q.H., Bai, H.R.*, et al.* (2021) Lysine crotonylation
873 of DgTIL1 at K72 modulates cold tolerance by enhancing DgnsLTP stability in chrysanthemum.
874 *Plant Biotechnology Journal* **19**, 1125-1140.
875 Huang, S.F., Kang, M.J. and Xu, A.L. (2017) HaploMerger2: rebuilding both haploid sub-assemblies from
876 high-heterozygosity diploid genome assembly. *Bioinformatics* **33**, 2577-2579.
877 Jahan, K., Nie, H. and Yan, X. (2023) Revealing the potential regulatory relationship between HSP70,
878 HSP90 and HSF genes under temperature stress. *Fish & Shellfish Immunology* **134**, 108607.
879 Jiang, H., Zhou, L.J., Gao, H.N., Wang, X.F., Li, Z.W. and Li, Y.Y. (2022) The transcription factor MdMYB2
880 influences cold tolerance and anthocyanin accumulation by activating SUMO E3 ligase MdSIZ1
881 in apple. *Plant Physiol* **189**, 2044-2060.
882 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. and Jermiin, L.S. (2017) ModelFinder:
883 fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587-589.
884 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*
885 **28**, 27-30.
886 Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7:
887 improvements in performance and usability. *Mol Biol Evol* **30**, 772-780.
888 Ke, L., Lei, W., Yang, W., Wang, J., Gao, J., Cheng, J., Sun, Y.*, et al.* (2020) Genome-wide identification of
889 cold responsive transcription factors in *Brassica napus* L. *BMC Plant Biol* **20**, 62.
890 Keilwagen, J., Hartung, F. and Grau, J. (2019) GeMoMa: Homology-Based Gene Prediction Utilizing
891 Intron Position Conservation and RNA-seq Data. *Methods Mol Biol* **1962**, 161-177.
892 Kidokoro, S., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2022) Transcriptional regulatory network of
893 plant cold-stress responses. *Trends Plant Sci* **27**, 922-935.
894 Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and
895 genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915.
896 Kokot, M., Dlugosz, M. and Deorowicz, S. (2017) KMC 3: counting and manipulating k-mer statistics.
897 *Bioinformatics* **33**, 2759-2761.
898 Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
899 applications. *Bioinformatics* **27**, 1571-1572.
900 Kupsch, C., Ruwe, H., Gusewski, S., Tillich, M., Small, I. and Schmitz-Linneweber, C. (2012) Arabidopsis
901 chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and
902 confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell*
903 **24**, 4266-4280.
904 Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T. and Ussery, D.W. (2007) RNAmmer:
905 consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100-3108.
906 Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis.
907 *BMC Bioinformatics* **9**, 559.
908 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-
909 359.
910 Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100.
911 Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.
912 *Bioinformatics* **25**, 1754-1760.
913 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.*, et al.* (2009) The Sequence

914 Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
915 Lin, K.H., Sei, S.C., Su, Y.H. and Chiang, C.M. (2019) Overexpression of the Arabidopsis and winter squash
916 superoxide dismutase genes enhances chilling tolerance via ABA-sensitive transcriptional
917 regulation in transgenic Arabidopsis. *Plant Signal Behav* **14**, 1685728.
918 Lin, Y., Ye, C., Li, X., Chen, Q., Wu, Y., Zhang, F., Pan, R.*, et al.* (2023) quarTeT: a telomere-to-telomere
919 toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res* **10**,
920 uhad127.
921 Liu, Z., Dong, X., Cao, X., Xu, C., Wei, J., Zhen, G., Wang, J.*, et al.* (2022) QTL mapping for cold tolerance
922 and higher overwintering survival rate in winter rapeseed (*Brassica napus*). *J Plant Physiol* **275**,
923 153735.
924 Lu, Y., Hu, Y. and Li, P. (2017) Consistency of electrical and physiological properties of tea leaves on
925 indicating critical cold temperature. *Biosystems Engineering* **159**, 89-96.
926 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.*, et al.* (2010)
927 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
928 sequencing data. *Genome Res* **20**, 1297-1303.
929 Mi, W., Liu, Z., Jin, J., Dong, X., Xu, C., Zou, Y., Xu, M.*, et al.* (2021) Comparative proteomics analysis
930 reveals the molecular mechanism of enhanced cold tolerance through ROS scavenging in
931 winter rapeseed (*Brassica napus* L.). *PLoS One* **16**, e0243292.
932 Nawrocki, E.P. and Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*
933 **29**, 2933-2935.
934 Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: a fast and effective
935 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-
936 274.
937 Orvar, B.L., Sangwan, V., Omann, F. and Dhindsa, R.S. (2000) Early steps in cold sensing by plant cells:
938 the role of actin cytoskeleton and membrane fluidity. *Plant J* **23**, 785-794.
939 Ou, S. and Jiang, N. (2018) LTR_retriever: A Highly Accurate and Sensitive Program for Identification of
940 Long Terminal Repeat Retrotransposons. *Plant Physiol* **176**, 1410-1422.
941 Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in
942 eukaryotic genomes. *Bioinformatics* **23**, 1061-1067.
943 Peng, H. and Zhang, J. (2009) Plant genomic DNA methylation in response to stresses: Potential
944 applications and challenges in plant breeding. *Progress in Natural Science* **19**, 1037-1045.
945 Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie
946 enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**,
947 290-295.
948 Probst, A.V. and Mittelsten Scheid, O. (2015) Stress-induced structural changes in plant chromatin. *Curr*
949 *Opin Plant Biol* **27**, 8-16.
950 Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-
951 redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-
952 65.
953 Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005)
954 InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-120.
955 Ranallo-Benavidez, T.R., Jaron, K.S. and Schatz, M.C. (2020) GenomeScope 2.0 and Smudgeplot for
956 reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432.
957 Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E.*, et al.* (2015) HiC-Pro: an

958 optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259.
959 Shi, Y., Ding, Y. and Yang, S. (2018) Molecular Regulation of CBF Signaling in Cold Acclimation. *Trends*
960 *Plant Sci* **23**, 623-637.
961 Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO:
962 assessing genome assembly and annotation completeness with single-copy orthologs.
963 *Bioinformatics* **31**, 3210-3212.
964 Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D.*, et al.* (2020) Eight high-quality genomes
965 reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat Plants* **6**,
966 34-45.
967 Stanke, M., Steinkamp, R., Waack, S. and Morgenstern, B. (2004) AUGUSTUS: a web server for gene
968 finding in eukaryotes. *Nucleic Acids Res* **32**, W309-312.
969 Su, C.F., Wang, Y.C., Hsieh, T.H., Lu, C.A., Tseng, T.H. and Yu, S.M. (2010) A novel MYBS3-dependent
970 pathway confers cold tolerance in rice. *Plant Physiol* **153**, 145-158.
971 Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., Tong, C., Li, J.*, et al.* (2017) The high-quality genome of
972 *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype.
973 *Plant J* **92**, 452-468.
974 Sun, H.Q., Ding, J., Piednoël, M. and Schneeberger, K. (2018) findGSE: estimating genome size variation
975 within human and Arabidopsis using k-mer frequencies. *Bioinformatics* **34**, 550-557.
976 Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M. and Paterson, A.H. (2008) Synteny and collinearity
977 in plant genomes. *Science* **320**, 486-488.
978 Tarailo-Graovac, M. and Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic
979 sequences. *Curr Protoc Bioinformatics* **Chapter 4**, 4 10 11-14 10 14.
980 Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. and Prins, P. (2015) Sambamba: fast processing of NGS
981 alignment formats. *Bioinformatics* **31**, 2032-2034.
982 Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B.*, et al.*
983 (2001) The COG database: new developments in phylogenetic classification of proteins from
984 complete genomes. *Nucleic Acids Res* **29**, 22-28.
985 Vannini, C., Locatelli, F., Bracale, M., Magnani, E., Marsoni, M., Osnato, M., Mattana, M.*, et al.* (2004)
986 Overexpression of the rice *Osmyb4* gene increases chilling and freezing tolerance of
987 Arabidopsis thaliana plants. *Plant J* **37**, 115-127.
988 Wang, X. and Wang, L. (2016) GMATA: An Integrated Software Package for Genome-Scale SSR Mining,
989 Marker Development and Viewing. *Front Plant Sci* **7**, 1350.
990 Wei, D., Cui, Y., He, Y., Xiong, Q., Qian, L., Tong, C., Lu, G.*, et al.* (2017) A genome-wide survey with
991 different rapeseed ecotypes uncovers footprints of domestication and breeding. *J Exp Bot* **68**,
992 4791-4801.
993 Wei, J., Shen, Y., Dong, X., Zhu, Y., Cui, J., Li, H., Zheng, G.*, et al.* (2022) DNA methylation affects freezing
994 tolerance in winter rapeseed by mediating the expression of genes related to JA and CK
995 pathways. *Front Genet* **13**, 968494.
996 Wei, J., Zheng, G., Yu, X., Liu, S., Dong, X., Cao, X., Fang, X.*, et al.* (2021) Comparative Transcriptomics
997 and Proteomics Analyses of Leaves Reveals a Freezing Stress-Responsive Molecular Network in
998 Winter Rapeseed (*Brassica rapa* L.). *Front Plant Sci* **12**, 664311.
999 Wingett, S.W. and Andrews, S. (2018) FastQ Screen: A tool for multi-genome mapping and quality
1000 control. *F1000Res* **7**, 1338.
1001 Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G.*, et al.* (2019) Whole-Genome Resequencing

1002 of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype
1003 Divergence. *Mol Plant* **12**, 30-43.
1004 Wu, L., Tian, J., Yu, Y., Yuan, L., Zhang, Y., Wu, H., Wang, F.*, et al.* (2023) Functional characterization of a
1005 cold related flavanone 3-hydroxylase from Tetrastigma hemsleyanum: an in vitro, in silico and
1006 in vivo study. *Biotechnol Lett*.
1007 Wu, R.G., Wang, Y., Wu, T., Xu, X.F. and Han, Z.H. (2018) Functional characterisation of *MdMYB44* as a
1008 negative regulator in the response to cold and salt stress in apple calli. *J Hortic Sci Biotech* **93**,
1009 347-355.
1010 Xie, Y., Chen, P., Yan, Y., Bao, C., Li, X., Wang, L., Shen, X.*, et al.* (2018) An atypical R2R3 MYB transcription
1011 factor increases cold hardiness by CBF-dependent and CBF-independent pathways in apple.
1012 *New Phytol* **218**, 201-218.
1013 Xu, H., Yang, G., Zhang, J., Wang, Y., Zhang, T., Wang, N., Jiang, S.*, et al.* (2018a) Overexpression of a
1014 repressor *MdMYB15L* negatively regulates anthocyanin and cold tolerance in red-fleshed callus.
1015 *Biochem Biophys Res Commun* **500**, 405-410.
1016 Xu, H.F., Wang, N., Wang, Y.C., Jiang, S.H., Fang, H.C., Zhang, J., Su, M.Y.*, et al.* (2018b) Overexpression
1017 of the transcription factor MdbHLH33 increases cold tolerance of transgenic apple callus. *Plant*
1018 *Cell Tiss Org* **134**, 131-140.
1019 Xuan, L., Yan, T., Lu, L., Zhao, X., Wu, D., Hua, S. and Jiang, L. (2020) Genome-wide association study
1020 reveals new genes involved in leaf trichome formation in polyploid oilseed rape (*Brassica napus*
1021 L.). *Plant Cell Environ* **43**, 675-691.
1022 Yang, A., Dai, X. and Zhang, W.H. (2012) A R2R3-type MYB gene, *OsMYB2*, is involved in salt, cold, and
1023 dehydration tolerance in rice. *J Exp Bot* **63**, 2541-2556.
1024 Ye, Y., Ding, Y., Jiang, Q., Wang, F., Sun, J. and Zhu, C. (2017) The role of receptor-like protein kinases
1025 (RLKs) in abiotic stress response in plants. *Plant Cell Rep* **36**, 235-242.
1026 Yu, J., Cang, J., Lu, Q., Fan, B., Xu, Q., Li, W. and Wang, X. (2020) ABA enhanced cold tolerance of wheat
1027 'dn1' via increasing ROS scavenging system. *Plant Signal Behav* **15**, 1780403.
1028 Yu, T., Zhang, J., Cao, J., Cai, Q., Li, X., Sun, Y., Li, S.*, et al.* (2021) Leaf transcriptomic response mediated
1029 by cold stress in two maize inbred lines with contrasting tolerance levels. *Genomics* **113**, 782-
1030 794.
1031 Yuan, Y., Bayer, P.E., Batley, J. and Edwards, D. (2021) Current status of structural variation studies in
1032 plants. *Plant Biotechnol J* **19**, 2153-2163.
1033 Zhao, Y., Antoniou-Kourounioti, R.L., Calder, G., Dean, C. and Howard, M. (2020) Temperature-
1034 dependent growth contributes to long-term cold sensing. *Nature* **583**, 825-829.
1035 Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies.
1036 *Nat Genet* **44**, 821-824.
1037
1038

Figure and table legends

Table 1. Summary of genome assemblies and annotation for NTS57 and other public rapeseed genomes.

Accessions	Assembly size (MB)	Anchored percentage (%)	Contig N50 (Mb)	No. of predicted genes	Repeats (%)	BUSCO (%)	No. of gaps	Reference
Darmor-*bzh* (v10)	924	93.84	11.48	108,190	53.86	98.60	268	Rousseau-Gueutin et al.
Westar	1007	93.54	3.13	97514	55.78	98.00	2164	Song et al.
No2127	1011	92.18	2.70	95385	56.86	98.00	2073	Song et al.
Zheyou7	1015	91.03	2.10	96209	56.39	98.30	3027	Song et al.
Gangan	1033	90.41	2.71	96843	56.58	98.30	1976	Song et al.
Shengli	1001	91.98	2.82	94586	56.29	98.10	2241	Song et al.
Tapidor	1014	91.91	2.85	96117	56.59	98.00	1871	Song et al.
Quinta	1003	93.02	2.80	95492	56.29	98.30	2132	Song et al.
ZS11	1010	95.33	2.70	100,919	55.62	98.10	5460	Song et al.
ZS11	921	93.10	1.64	106,059	61.83	98.30	361	Chen et al.
NTS57	1062	93.60	21.81	101,438	55.60	99.75	66	This study

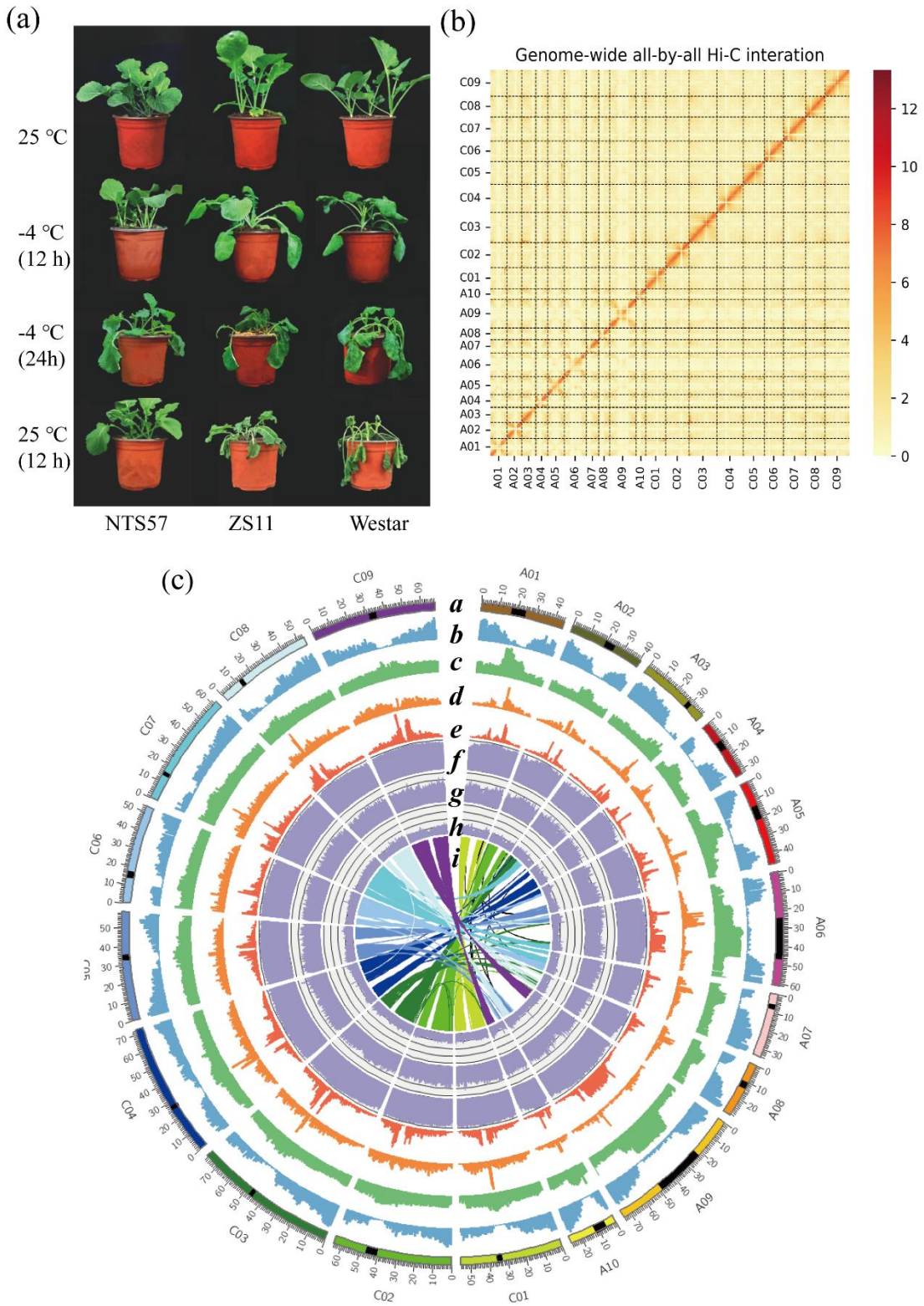
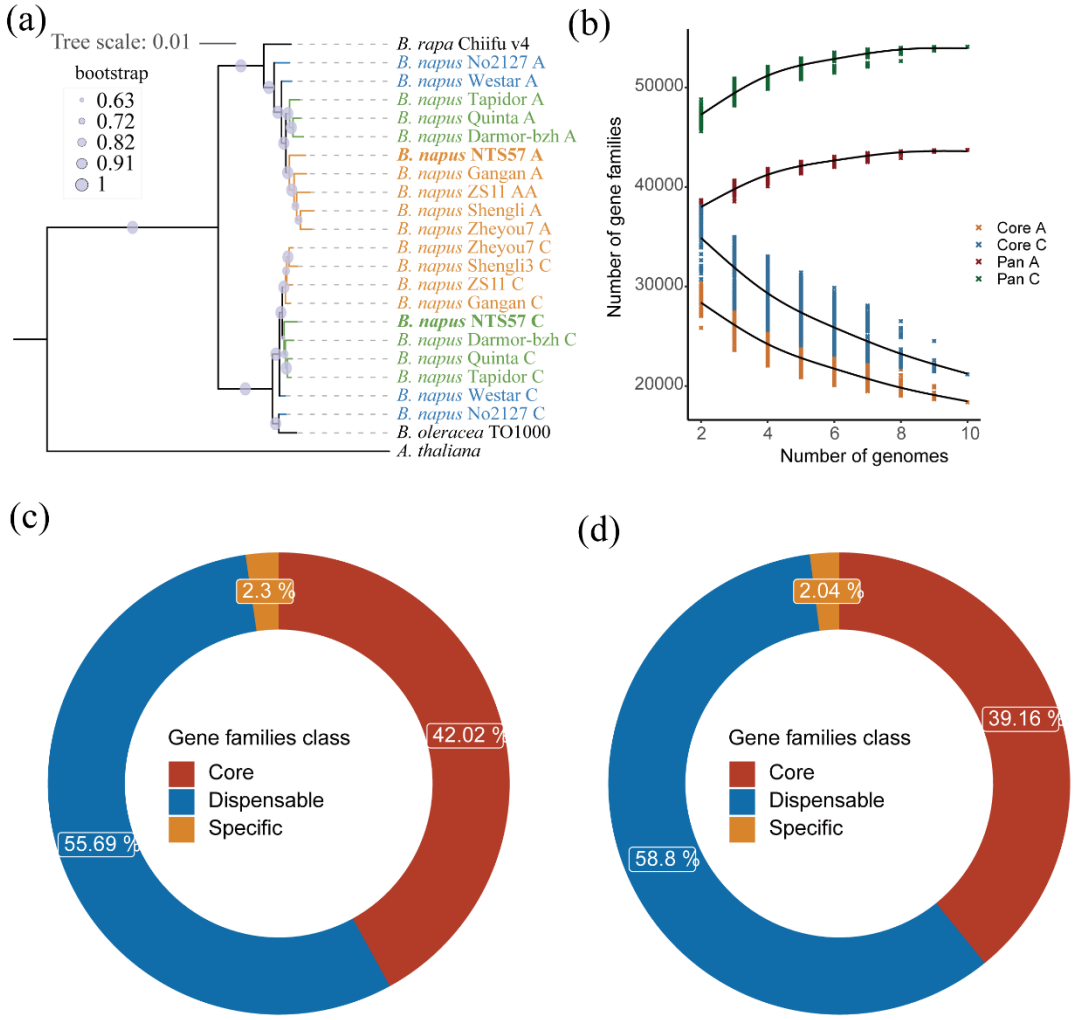


Figure 1. Cold tolerance phenotypes and genome assembly of NTS57. (a) Seedling survival of NTS57, ZS11 and Westar. Seedlings were photographed before or 12h and 24h after treatment at -4 °C, and 12h after returning to normal conditions at 25 °C. **(b)** Circos plot showing the genome assembly and genomic features of NTS57. The letters

1049 (*a-i*) in the circos indicate the ideograms of 19 chromosomes of the NTS57 genome,
1050 gene density, GC content, Copia transposons, Gypsy transposons, CpG methylation,
1051 CHG methylation, CHH methylation, and chromosomal collinearity, respectively. The
1052 centromeres are highlighted in black in the ideograms. **(c)** Genome-wide Hi-C heat map
1053 of the NTS57 genome. The heatmap shows the diagonal pattern of strong Hi-C intra-
1054 chromosomal interactions of NTS57. The color bar indicates the interaction frequencies
1055 of the Hi-C chromosome links.



1056

1057 **Figure 2. Phylogenetic relationships and pan-genomic analysis of NTS57 and other**
1058 **rapeseed accessions. (a)** Maximum likelihood tree constructed using 2,268 single-
1059 copy genes showing the phylogenetic relationships of NTS57 and other public rapeseed
1060 accessions. Bootstrap values are given as percentages on the branches of the consensus
1061 tree inferred from 1000 replicates. **(b)** Modelling of pangenome and coregenome sizes
1062 by including additional rapeseed accessions. **(c)** and **(d)** show the percentage of core

genes, dispensable genes and accession-specific genes in the A and C subgenomes of rapeseed, respectively.

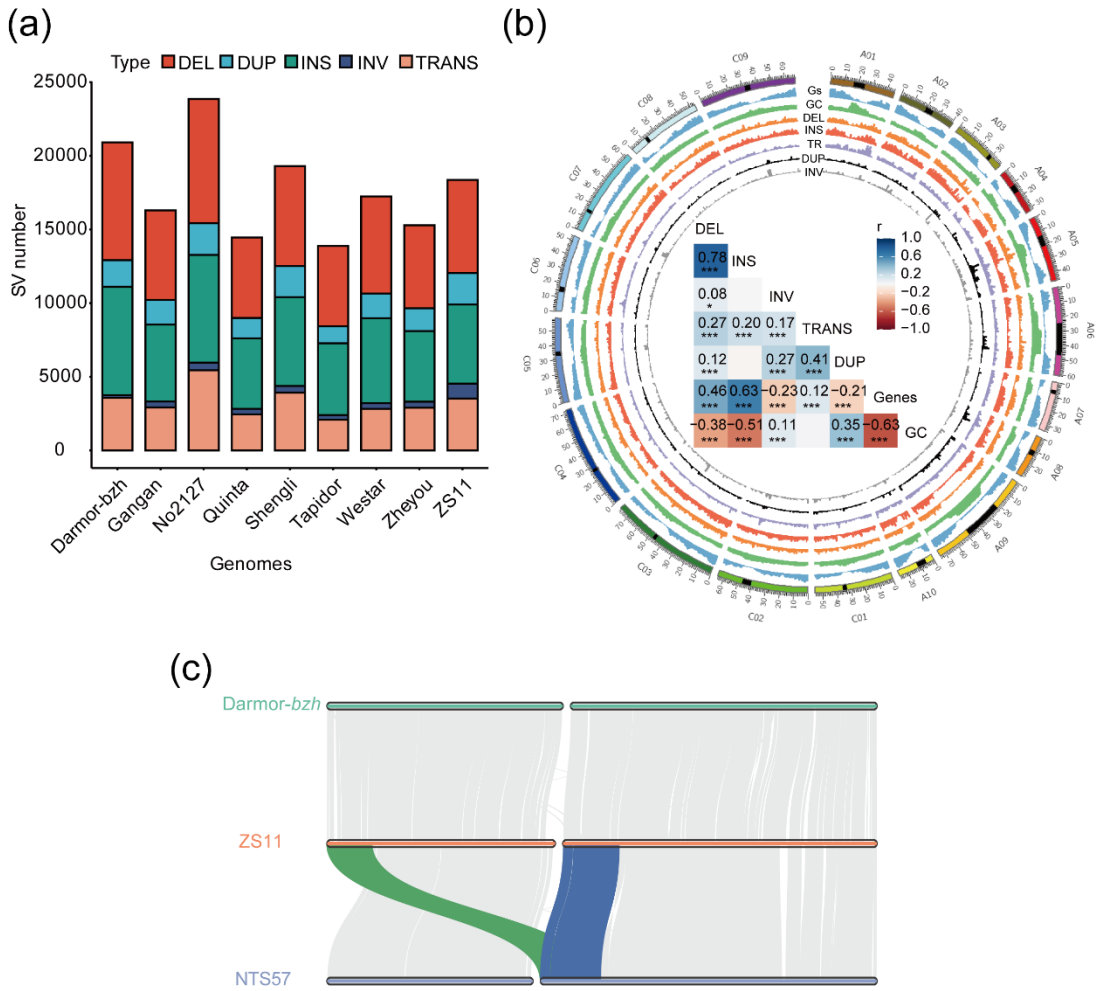


Figure 3. Genome structure variation between NTS57 and other rapeseed accessions. (a) Number of different SV types within each rapeseed accession using NTS57 as the reference genome. “TRANS”, “INV”, “INS”, “DUP”, and “DEL” stand for translocation, inversion, insertion, duplication, and deletion, respectively. **(b)** SV landscape in the NTS57 genome. The heat map shows the correlations between different SVs, GC content and gene density, with the correlation significance was marked in the corresponding cells. **(c)** Chromosomal synteny for A03 and C03 between Darmor-*bzh*, ZS11 and NTS57. The green and blue ribbons show the synteny blocks of the leftmost region (~6.6 Mb) of C03 in NTS57 with that of A03 and C03 of ZS11, respectively.

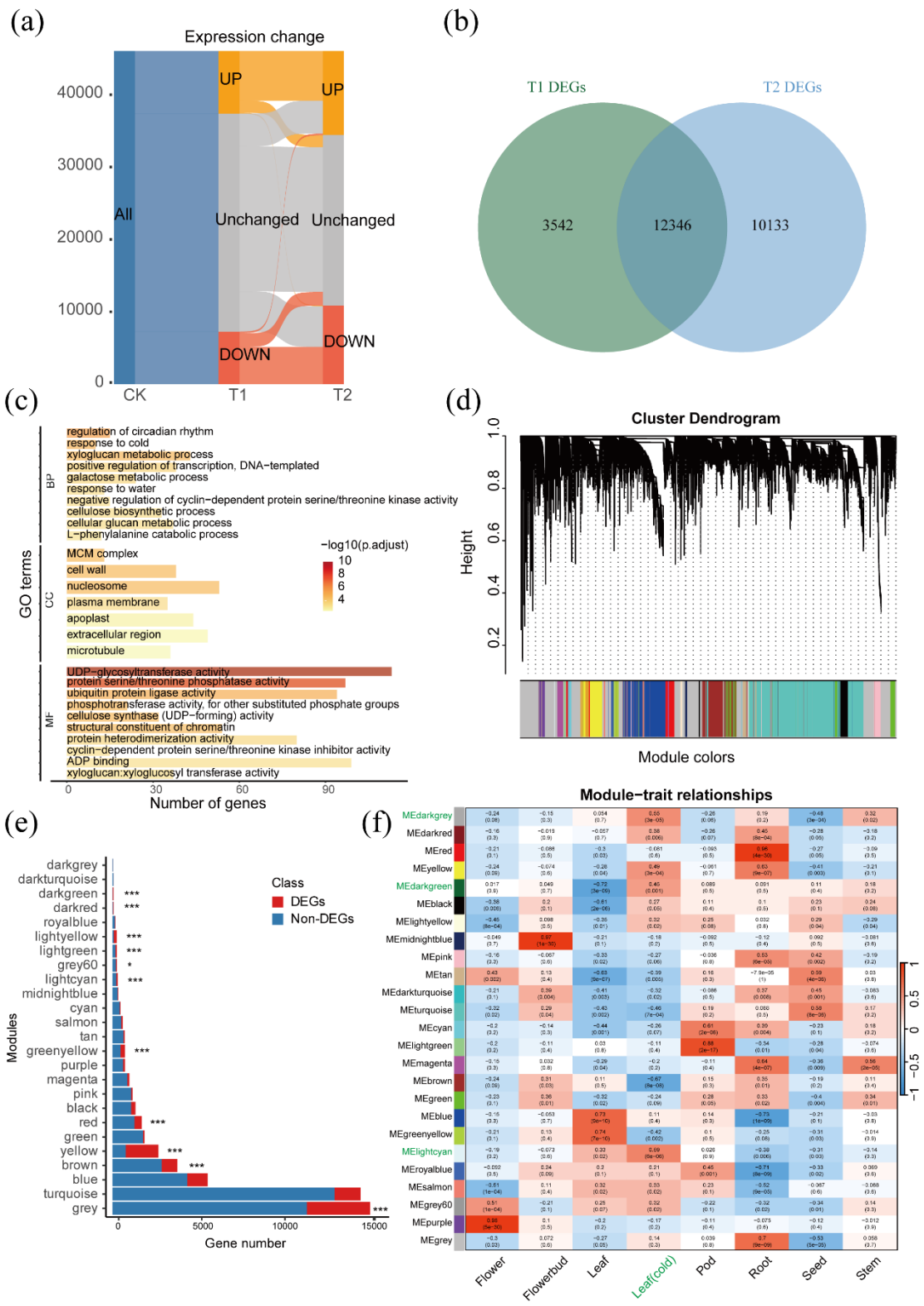


Figure 4. Gene expression analysis of NTS57 under cold stress. (a) Sankey plot showing the up- and down-regulated genes of NTS57 after 12h (T1) and 24h (T2) cold stress. **(b)** Shared DEGs in T1 and T2 samples. **(c)** Gene function enrichment of DEGs in T1 samples. **(d)** Co-expression gene modules identified by WGCNA analysis, with

different colors denoting the different gene modules. **(e)** Enrichment of DEGs in each of the co-expression gene modules. **(f)** Heat map showing the module-trait relationships. The color bar denotes the correlation of the modules with each trait.

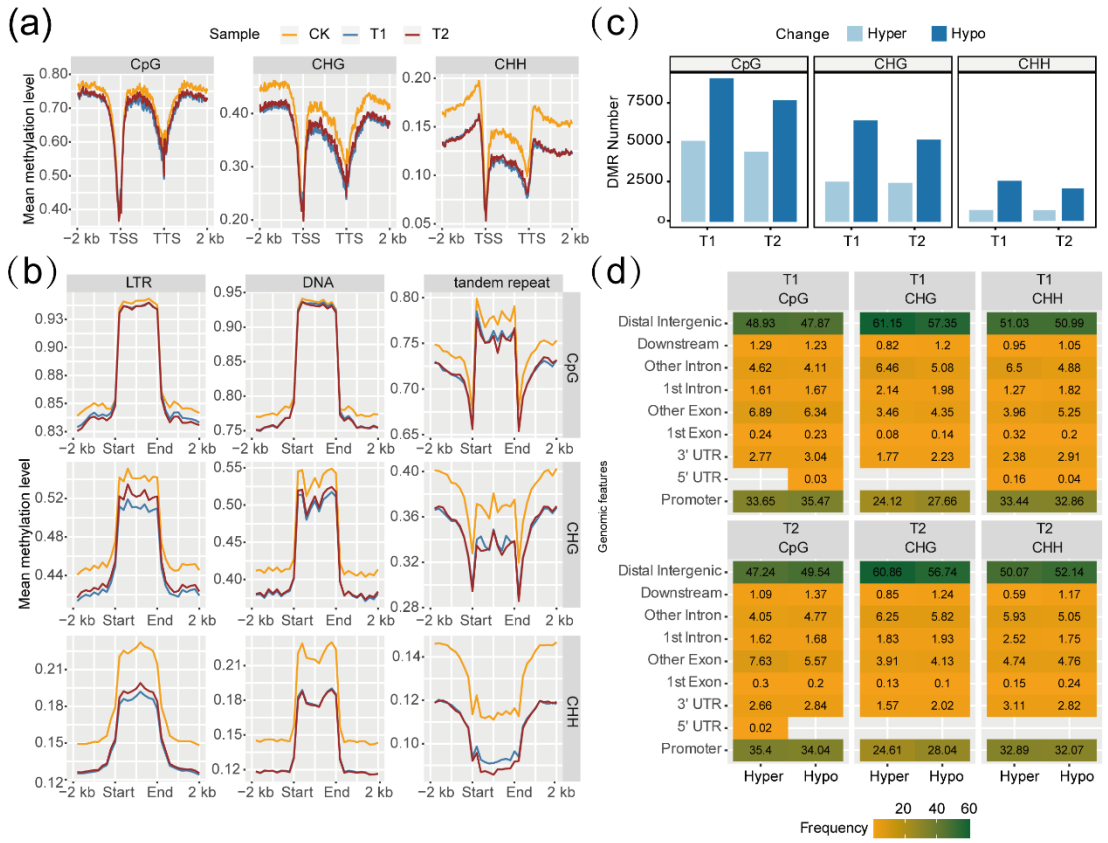


Figure 5. Methylation dynamics of NTS57 under cold stress. (a) Profiles of methylation levels in the gene body, upstream and downstream regions in the NTS57 genome under normal and cold stress. The gene body was divided into 100 bins of equal size. For each bin in each gene, the methylation level was calculated by dividing the total number of methylated Cs by the total number of Cs, and the average of methylation level in the window across all protein genes was used to generate the profiles. The three methylation profiles for CpG, CHG and CHH contexts were analyzed independently. **(b)** Similar to (a) but showing the profiles of methylation levels in the repetitive elements of TNS57. T1 and T2 represent 12 and 24 h after cold treatment, and the CK represents the normal condition. **(c)** Number of differentially methylated regions (DRMs) in T1 and T2 samples relative to CK. DMRs were classified into hypermethylation (hyper) and hypomethylation (hypo). **(d)** The percentage of DMRs in different genomic contexts.

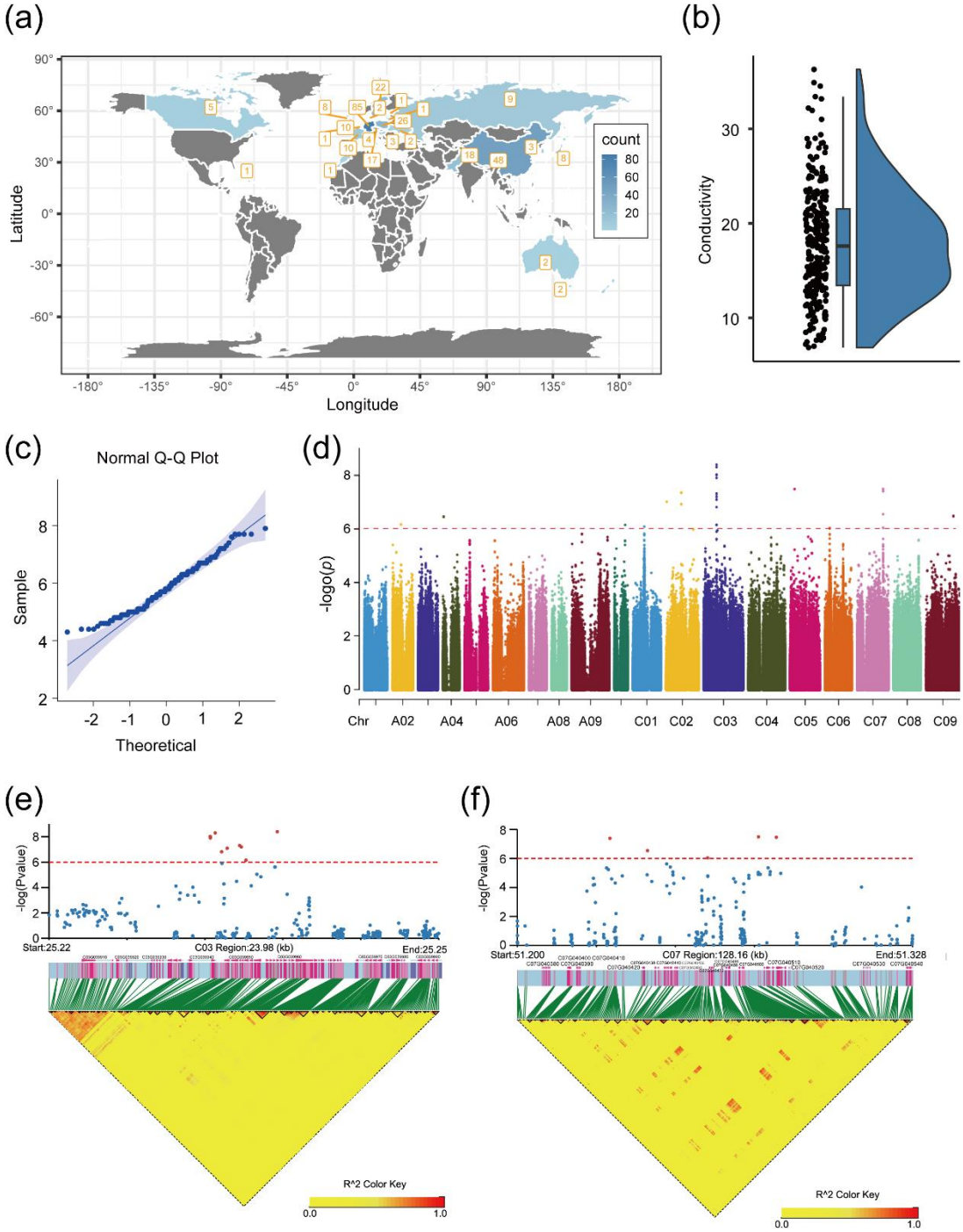


Figure 6. GWAS identify association signals for cold tolerance in rapeseeds. (a) Geographical distribution of the 290 rapeseed accessions used in the GWAS analysis. **(b)** Distribution of relative electrical conductivity for the 290 rapeseed accessions. **(c)** QQ plot showing the normal distribution of the relative electrical conductivity for the 290 rapeseed accessions. **(d)** Manhattan plot of GWAS analysis for relative electrical conductivity. **(e)** Local Manhattan plot. Gene positions and LD heat map show the

regions surrounding the lead SNPs and the associated candidate genes by SNP-GWAS analysis. **(f)** Same as **(e)** but a different GWAS block on chromosome C07.

Supporting Information

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

Figure S1. Genome size estimation based on survey sequencing.

Figure S2. Bias in gene expression between two sub-genomes of NTS57.

Figure S3. Venn diagram showing the results of gene function annotations using different databases.

Figure S4. Gene function enrichment for core genes and dispensable genes of NTS57.

Figure S5. Comparisons of gene expression levels for three classes of genes in NTS57.

Figure S6. Length statistics for structural variations in different rapeseed accessions using NTS57 as the reference genome.

Figure S7. Number of structural variations in different rapeseed accessions using ZS11 as the reference genome.

Figure S8. Length statistics for structural variation in different rapeseed accessions using ZS11 as the reference genome.

Figure S9. Synteny and rearrangement (SyRi) plot for 19 chromosomes of NTS57 and ZS11.

Figure S10. PacBio long reads mapped at the end of the genomic region of NTS57 that had a syntenic relationship to A03 of the ZS11 genome.

Figure S11. Reproducibility of differentially expressed genes identified by different programs.

Figure S12. Gene function enrichment for the DEGs identified in T2 samples relative to CK.

Figure S13. KEGG pathway enrichment for the DEGs identified in T1 or T2

1134 **samples relative to CK.**

1135 **Figure S14. RNA-Seq samples used to construct the GWCNA gene co-expression**
1136 **network.**

1137 **Figure S15. Soft-thresholding power determination in WGCNA.**

1138 **Figure S16. Heat map showing the expression profiles of the genes in the brown**
1139 **module from WGCNA.**

1140 **Figure S17. Heat map showing the expression profiles of the genes in the**
1141 **darkgreen module from WGCNA.**

1142 **Figure S18. Heat map showing the expression profiles of the genes in the lightcyan**
1143 **module from WGCNA.**

1144 **Figure S19. Enrichment of gene functions for three focused modules.**

1145 **Figure S20. Methylation level profiles of A- and C-subgenomes of NTS57 under**
1146 **cold stress and normal conditions.**

1147 **Figure S21. Number of DMR-associated genes in T1 and T2 samples.**

1148 **Figure S22. Number of DMR-associated DEGs in T1 and T2 samples.**

1149 **Figure S23. Number of up- and down-regulated DMR-associated DEGs in T1 and**
1150 **T2 samples.**

1151 **Figure S24. Histogram showing the number of rapeseed accessions with different**
1152 **eco-types that collected from 24 countries.**

1153 **Figure S25. Chromosomal distribution of the 9,446,761 SNPs identified in 289**
1154 **rapeseed accessions.**

1155 **Figure S26. Chromosomal distribution of the 469,176 InDels identified in 289**
1156 **rapeseed accessions.**

1157 **Figure S27. PCA analysis of 289 rapeseed accessions based on SNPs (a) and InDels**
1158 **(b).**

1159 **Figure S28. Manhattan plot of the InDel-GWAS analysis for relative electronic**
1160 **conductivity.**

1161 **Figure S29. Effect of the lead SNPs on relative electronic conductivity on the**
1162 **chromosome C03 of the NTS57 genome.**

1163 **Figure S30. Effect of the lead SNPs on relative electronics conductivity on the**

1164 **chromosome A07 of the NTS57 genome.**

1165 **Figure S31. Genome Browser snapshot showing the expression change of three**
1166 **genes associated with the lead SNPs from the GWAS analysis.**

1167 **Figure S32 qRT-PCR analysis of two genes associated with cold tolerance in**
1168 **rapeseed.**

1169 **Table S1. Distribution of predicted centromeres in the NTS57 reference genome.**

1170 **Table S2. Comparison of gene number between different rapeseed accessions.**

1171 **Table S3. Summary of non-coding RNAs in the NTS57 genome.**

1172 **Table S4. Summary of repetitive elements in the NTS57 genome.**

1173 **Table S5. Genes associated with SVs on the chromosome C03 of NTS57**

1174 **Table S6. Summary of differentially expressed genes after 12 hours of cold**
1175 **exposure.**

1176 **Table S7. Summary of differentially expressed genes after 24 hours of cold**
1177 **exposure.**

1178 **Table S8. NTS57 specific genes that were differentially expressed in cold stress.**

1179 **Table S9. Gene expression levels of NTS57 across different developmental stages**
1180 **and tissues.**

1181 **Table S10. Summary of the significant SNPs associated with cold tolerance of**
1182 **rapeseeds.**

1183 **Table S11. Summary of the significant InDels associated with cold tolerance of**
1184 **rapeseeds.**

1185 **Table S12. List of primers used for RT-qPCR in this study.**

1186