Genome-wide dissection of seed glucosinolate content identifies candidate genes and favorable haplotypes for quality breeding in Brassica napus L.

Kaijie Ye
Southwest University
Liang Ran
Southwest University
Qiuting Wei
Southwest University
Nengwen Yin
Southwest University
Huiyan Zhao
Southwest University
Ti Zhang
Southwest University
Taocui Huang
Chongqing Academy of Agricultural Sciences
Hai Du
Southwest University
Cunmin Qu
Southwest University
Haijiang Liu
Southwest University
Huafang Wan
wanhua05@swu.edu.cn
Southwest University

Research Article

Keywords: Oilseed rape, genome-wide association study, seed glucosinolate content, candidate genes, favorable haplotype

Posted Date: June 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9844632/v1

Additional Declarations: No competing interests reported.

Abstract

Seed glucosinolate content (SGC) is a key quality trait in rapeseed (*Brassica napus* L.) that directly affects meal utilization and breeding value. Here, we evaluated SGC in a natural population of 403 *B. napus* accessions across three environments and performed genome-wide association analysis using 1.60 million high-quality SNPs. SGC displayed extensive phenotypic variation, strong correlations across environments, and high broad-sense heritability, indicating a strong genetic basis. A total of 11 significant QTLs associated with SGC were identified on chromosomes A01, A02, A03, A08, A09, C02, C03, C07, C08, and C09, among which several were repeatedly detected across environments and in BLUP data. Integration of stable GWAS loci with seed expression–phenotype association analysis based on the BnIR database prioritized four candidate genes: *BnaA03g40240D*, *BnaA09g05670D*, *BnaC02g42260D*, and *BnaC07g31100D*. Candidate-gene association and haplotype analyses identified favorable alleles and haplotypes conferring reduced SGC. Notably, pyramiding these favorable haplotypes reduced SGC by 70.2% and was simultaneously associated with higher seed oil content, improved unsaturated fatty acid composition, and markedly lower erucic acid content. Importantly, these quality improvements were achieved without significant effects on yield-related traits or root architecture. Collectively, this study provides new insights into the genetic architecture of SGC and valuable molecular resources for breeding low-glucosinolate, high-quality rapeseed cultivars.

1.Introduction

Vegetable oils are essential to human nutrition and play critical roles in food, industrial, and pharmaceutical applications (Voon et al. 2024). As the world's third-largest oilseed crop, rapeseed (*Brassica napus* L.) accounts for approximately 13% of global vegetable oil production. (USDA ERS, 2022). Developing high-quality rapeseed varieties remains a primary breeding goal, with seed quality predominantly determined by seed oil content (SOC), seed glucosinolate content (SGC) and erucic acid. (Chao et al. 2022; Liu et al. 2025; Zhao et al. 2022)

During oil extraction, glucosinolates are hydrolyzed into isothiocyanates and other breakdown products, which can impair human health by disrupting thyroid function when consumed in excess (Tan et al. 2022). Moreover, high-SGC rapeseed meal, a major protein-rich byproduct, induces goiter and growth retardation in livestock, thereby limiting its feed value. Consequently, breeding low-SGC rapeseed cultivars is essential for enhancing human nutritional safety and improving the economic viability of rapeseed production. Traditionally, rapeseed improvement has relied on crossbreeding and mutagenesis, which have successfully modified various seed quality traits (Hassan 2014; Siddiqui et al. 2009). For example, both chemical and physical mutagenesis strategies have been employed to enhance agronomic traits., including pod size, seed weight, and oil content. (Siddiqui et al. 2009). However, these conventional approaches are often constrained by lengthy breeding cycles, random mutational events, and limited precision (Hu et al. 2021). Therefore, pinpointing the genetic loci underlying SGC and characterizing favorable haplotypes represent crucial prerequisites for accelerating marker-assisted and molecular breeding in *B. napus*.

Dissecting the genetic architecture of seed quality traits in *B. napus* has traditionally relied on linkage mapping and association analysis. For instance, linkage mapping in doubled haploid (DH) populations has successfully identified stable quantitative trait loci (QTLs) for SGC, such as *qGSL-C2*, which explains 30.88–72.87% of phenotypic variance across multiple environments (Liu et al. 2020). Similarly, 59 SGC-associated QTLs were detected (Chao et al. 2022) in a DH population evaluated across fifteen environments and, through integrated transcriptome profiling, proposed four candidate genes (*BnaC09.MYB28*, *BnaC02.GTR2a*, *BnaC09.SUR1*, and *BnaA09.APK1*). Despite these advances, QTL intervals often span several megabases and encompass hundreds of genes, complicating the precise identification of causal variants (Xiao et al. 2017). Following the advent of high-throughput sequencing, genome-wide association studies (GWAS) have emerged as a potent alternative, achieving greater resolution by exploiting historical recombination events within natural populations. GWAS has been successfully employed to pinpoint genes underlying complex traits. For example, GWAS identified BnaA09.MRP5 as a key regulator controlling seed phytic acid content (Liu et al. 2021). In recent years, performing GWAS across multiple environments has emerged as a highly efficient strategy for detecting stable QTL and refining candidate genes. Tang et al. (2021) applied GWAS to dissect SOC, co-localizing a major QTL on chromosome A05 and identifying *BnaA05.PMT6* as a putative regulator; CRISPR/Cas9-mediated knockout of this gene elevated seed oil content by 8% over wild-type levels(Tang et al. 2021). Collectively, these studies highlight how multi-environment GWAS can accelerate gene isolation and functional characterization in rapeseed.

Despite the progress in mapping SGC-related loci, the systematic utilization of multi-environment GWAS to dissect the genetic basis of SGC in *B. napus* natural populations remains underexplored. To address this gap, we phenotyped SGC in a diverse *B. napus* association panel at three locations and performed GWAS. The objectives of this study are as follows: (1) pinpoint reliable genetic loci and high-confidence candidate genes linked to SGC; (2) validate co-localized genomic regions through multi-environment comparison; and (3) define superior alleles and haplotypes suitable for deployment in marker-assisted selection and low-SGC *B. napus* breeding. These results offer useful genetic materials and practical targets for precision breeding strategies designed to enhance seed quality in rapeseed.

2. Materials and methods

2.1 Plant materials, field trials, and phenotypic evaluation

A natural population consisting of 403 diverse *Brassica napus* accessions was used in this study, including 336 semi-winter, 47 spring, 13 winter, and 8 unclassified accessions. These accessions were collected from major breeding programs worldwide, including 355 from China, 22 from Europe, 8 from Japan, 4 from Canada, 5 from Australia, 2 from Korea, and 8 of unknown origin(Liu et al. 2021).

The association panel was grown in three environments, designated SS, WH, and XM. Field experiments in each environment were conducted using a randomized complete block design with three biological replicates. Each plot consisted of three rows, with eight plants per row, and all accessions were

managed according to standard local agronomic practices. At maturity, seeds were harvested, dried, and used for seed glucosinolate content (SGC) determination. SGC was measured by near-infrared spectroscopy (NIRS).

Descriptive statistics, including mean, minimum, maximum, range, skewness, kurtosis, and coefficient of variation (CV), were calculated using the psych package in R. Best linear unbiased prediction (BLUP) values across environments were estimated and used as integrated phenotypic values for subsequent analyses. Pearson's correlation coefficients among environments and BLUP values were calculated in R. Broad-sense heritability $\left(H^2\right)$ was estimated based on phenotypic data across multiple environments.

2.2 Genome-wide association analysis and BnIR-based candidate gene prioritization

Genotypic data for the association panel were obtained from a high-density SNP dataset containing 1.60 million high-quality markers. SNPs were retained after filtering with a minor allele frequency (MAF) > 0.05 and a missing rate < 0.20 (Tang et al. 2021). Population structure was inferred using ADMIXTURE, and the kinship matrix was calculated using FaST-LMM. Based on cross-validation error and kinship relationships, the panel was divided into five subpopulations, as described previously (Liu et al. 2021). Genome-wide association studies (GWAS) for SGC were conducted separately using phenotypic data from SS, WH, XM, and BLUP values with the FaST-LMM algorithm, which accounts for both population structure and kinship. A significance threshold of $P < 1.0 \times 10^{-10}$ was used to identify significant marker–trait associations. Manhattan plots and quantile–quantile (Q–Q) plots were generated using the CMplot package in R. Significant SNPs clustered on the same chromosome were considered to define one QTL.

The linkage disequilibrium (LD) decay distance of this population was previously estimated to be approximately 238 kb (Liu et al. 2023). Therefore, genomic regions extending ± 300 kb around each lead SNP were defined as candidate intervals for candidate gene mining. Stable QTLs repeatedly detected across multiple environments and/or in the BLUP dataset were selected for further analysis.

To prioritize candidate genes within stable GWAS loci, expression data were retrieved from the gene expression module of the BnIR database (https://yanglab.hzau.edu.cn/). Seed transcript abundance at 20 and 40 days after flowering (DAF) was examined across the association panel, and Pearson's correlation analysis was performed between gene expression levels and SGC. Genes located within stable QTL intervals and showing significant expression–phenotype correlations were considered high-confidence candidate genes.

2.3 Candidate gene association and haplotype analysis

Sequence variants for the prioritized candidate genes, including *BnaA03g40240D*, *BnaA09g05670D*, *BnaC02g42260D*, and *BnaC07g31100D*, were extracted from the population VCF file using VCFtools

v0.1.16 (Danecek et al. 2011). Variants located within the coding regions and the 2-kb upstream promoter regions of each gene were retained for further analysis.

Candidate gene-based association analysis was performed using TASSEL v5.0 (Bradbury et al. 2007). SNPs significantly associated with SGC were identified and used for subsequent haplotype analysis. Pairwise linkage disequilibrium (LD) and haplotype structure were analyzed using Haploview 4.2 (Barrett et al. 2005). Haplotypes represented by fewer than 10 accessions were excluded from downstream analyses.

Phenotypic differences among haplotypes were evaluated for SGC. For genes with multiple significant SNPs, major haplotypes were defined based on representative associated variants, and SGC differences among haplotypes were compared using one-way analysis of variance (ANOVA), followed by post hoc multiple comparison tests. For loci represented by a single significant SNP, phenotypic differences between alleles were compared directly.

2.4 Evaluation of integrated low-SGC haplotypes and associated traits

To assess the breeding potential of favorable haplotypes, accessions carrying the favorable low-SGC haplotypes identified from the four candidate genes were classified as the LSGC-hap group, whereas accessions carrying the corresponding unfavorable high-SGC haplotypes were classified as the HSGC-hap =group. Based on this criterion, 38 accessions were assigned to the LSGC-hap group and 57 accessions to the HSGC-hap group.

Phenotypic comparisons between the two groups were conducted for seed quality traits, agronomic traits, and root architecture traits. Seed quality traits included SGC, seed oil content (SOC), oleic acid, linoleic acid, linolenic acid, and erucic acid contents. Agronomic traits included seed yield (SY), seed weight, plant height, and branch number. Root traits included lateral root density (LRD), lateral root length (LRL), lateral root number (LRN), mean lateral root length (MLRL), primary root length (PRL), and total root length (TRL). Seed yield and other agronomic trait data were obtained from previous field experiments (Liu et al. 2022a; Liu et al. 2022b), and root trait data were obtained from our previous study (Wang et al. 2017). Differences between LSGC-hap and HSGC-hap groups were evaluated using two-tailed Student's *t*-tests in R. A probability level of $P < 0.05$ was considered statistically significant unless otherwise stated.

3. Results

3.1 Phenotypic Diversity and GWAS-Based Dissection of SGC

Seed glucosinolate content (SGC) was evaluated in a natural population of *B. napus* across three environments, including SS, WH, and XM. Substantial phenotypic variation was observed among

accessions in all environments (Fig. 1A). The mean SGC values were 33.27, 54.68, and 29.05 μmol/g in SS, WH, and XM, respectively, while the BLUP value showed an average of 40.10 μmol/g. Among the three environments, WH exhibited the highest mean SGC and the widest phenotypic range, with values varying from 0.01 to 187.07 μmol/g. In contrast, XM showed the lowest mean SGC, ranging from 0.01 to 155.18 μmol/g (Fig. 1A). High coefficients of variation were observed across environments, ranging from 83.43% in WH to 111.50% in SS, indicating extensive natural variation in SGC within the population (Fig. 1A).

Pairwise correlation analysis revealed strong positive correlations among SGC values measured in different environments and BLUP values (Fig. 1B). The Pearson correlation coefficients ranged from 0.86 to 0.97, suggesting that the phenotypic performance of accessions was largely consistent across environments. The correlations between individual environments and BLUP values were particularly high, with correlation coefficients of 0.94, 0.97, and 0.96 for XM, WH, and SS, respectively. Moreover, the broad-sense heritability of SGC was estimated to be 94.73% (Fig. 1B), indicating that SGC is a highly heritable trait mainly controlled by genetic factors. These results demonstrate that this natural population is well suited for genetic dissection of SGC. Moreover, the broad-sense heritability of SGC was estimated to be 94.73% (Fig. 1C), indicating that SGC is a highly heritable trait mainly controlled by genetic factors. These results demonstrate that this natural population is well suited for genetic dissection of SGC.

Using the FaST-LMM algorithm, GWAS was performed for SGC in the three environments and using BLUP values. Multiple significant association signals were detected across the *B. napus* genome (Fig. 2). In total, 11 significant QTLs were identified and distributed on chromosomes A01, A02, A03, A08, A09, C02, C03, C07, C08, and C09. Several loci were repeatedly detected across different environments and in the BLUP dataset, suggesting their stable effects on SGC variation. Notably, strong and consistent association peaks were observed on chromosomes A08 and C09, indicating that these regions may contain major loci regulating SGC accumulation. Together, these GWAS results provide a genetic basis for further candidate gene identification and functional analysis of SGC regulation in *B. napus*.

3.2 Identification of candidate genes through integration of stable GWAS loci and BnIR expression–phenotype associations

To further prioritize candidate genes underlying SGC variation, we focused on stable GWAS loci repeatedly detected across environments and/or in the BLUP dataset. Candidate intervals were defined as ± 300 kb flanking each lead SNP, according to the population LD decay distance of approximately 238 kb. Within these stable QTL intervals, a total of 108 positional candidate genes were retrieved for further analysis.

To identify genes with potential regulatory roles in SGC accumulation, we used the gene expression module of the BnIR database to examine the association between seed transcript abundance and SGC

across the natural population. Genes showing significant correlations between expression levels and SGC at seed developmental stages were prioritized as candidate regulators. Based on this strategy, four genes, *BnaA03g40240D*, *BnaA09g05670D*, *BnaC02g42260D*, and *BnaC07g31100D*, were selected as high-confidence candidates within stable GWAS loci (Fig. 3). The expression–phenotype correlation analysis revealed distinct stage-dependent patterns among these genes. At 20 DAF, *BnaA09g05670D*, *BnaC02g42260D*, and *BnaC07g31100D* showed significant positive correlations with SGC, with Pearson's correlation coefficients of 0.321, 0.431, and 0.358, respectively (Fig. 3B–D). Among them, *BnaC02g42260D* exhibited the strongest positive association with SGC at 20 DAF, suggesting a potentially important role during early seed development. In contrast, *BnaA03g40240D* showed no significant correlation with SGC at 20 DAF (R = − 0.005, P = 0.331; Fig. 3A).

At 40 DAF, *BnaA03g40240D* displayed a significant negative correlation with SGC (R = − 0.183, $P < 1.19 \times 10^{-8}$; Fig. 3E), indicating that it may act at a later seed developmental stage. *BnaA09g05670D* and *BnaC02g42260D* remained positively correlated with SGC at 40 DAF, with correlation coefficients of 0.276 and 0.368, respectively (Fig. 3F, G). By contrast, *BnaC07g31100D* showed a weak negative correlation with SGC at 40 DAF (R = − 0.0538, $P < 0.0041$; Fig. 3H), suggesting possible stage-specific or context-dependent regulation.

Together, these results indicate that the integration of stable GWAS loci with BnIR-based expression–phenotype association analysis effectively narrowed the candidate gene list and prioritized four genes potentially involved in the regulation of seed glucosinolate accumulation in *B. napus*.

3.3 Candidate-gene association and haplotype analyses identify favorable alleles for low-SGC breeding

To further dissect the allelic effects of the four prioritized candidate genes, we performed candidate-gene association analysis and haplotype analysis using SNPs located in the 2-kb upstream promoter regions and coding sequences of *BnaA09g05670D*, *BnaA03g40240D*, *BnaC02g42260D*, and *BnaC07g31100D*. Significant variants were then used to define major haplotypes and evaluate their effects on SGC.

For *BnaA09g05670D*, multiple SNPs within the gene region showed significant associations with SGC, including variants located in both regulatory and coding regions (Fig. 4A; Table 1). Four lead SNPs were selected for haplotype construction, resulting in four major haplotypes (Fig. 4B). Among them, *BnaA09g05670D*-Hap1 was the predominant haplotype, represented by 200 accessions, and exhibited the lowest mean SGC of 39.83 μmol/g. In contrast, Hap2, Hap3, and Hap4 showed substantially higher SGC levels, with mean values of 118.58, 98.69, and 120.04 μmol/g, respectively. Boxplot comparison further confirmed that Hap1 was significantly associated with reduced SGC compared with the other haplotypes (Fig. 4C). Therefore, *BnaA09g05670D*-Hap1 was identified as a favorable haplotype for low-SGC breeding.

Table 1
List of SNP variants identified in the candidate genes of *BnaA09g05670D*, *BnaA03g40240D*, *BnaC02g42260D* and *BnaC07g31100D*

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
BnaA09g05670D	chrA09__2774627	Exon	7.35E-19	20.86%	synonymous_variant
	chrA09__2774629	Exon	8.73E-19	20.85%	synonymous_variant
	chrA09__2774660	Exon	8.29E-24	25.82%	missense_variant
	chrA09__2774665	Exon	1.30E-20	22.54%	synonymous_variant
	chrA09__2774666	Exon	6.32E-20	21.76%	synonymous_variant
	chrA09__2774717	Exon	6.27E-23	23.04%	synonymous_variant
	chrA09__2774779	Exon	5.13E-12	13.61%	missense_variant
	chrA09__2774783	Exon	4.24E-12	13.67%	synonymous_variant
	chrA09__2774828	Exon	1.15E-06	7.52%	synonymous_variant
	chrA09__2774906	Exon	3.85E-13	14.83%	synonymous_variant
	chrA09__2774969	Exon	1.69E-09	11.33%	synonymous_variant
	chrA09__2775106	Exon	3.31E-18	21.46%	missense_variant
	chrA09__2775179	Exon	3.40E-20	22.40%	synonymous_variant
	chrA09__2775244	Exon	2.85E-15	17.64%	missense_variant
BnaA03g40240D	chrA03__20115975	Exon	1.39E-10	10.89%	synonymous_variant
	chrA03__20116017	Exon	0.08081	1.52%	synonymous_variant
	chrA03__20116023	Exon	2.78E-09	9.31%	synonymous_variant

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrA03__20116056	Exon	8.38E-10	9.75%	synonymous_variant
	chrA03__20116068	Exon	6.18E-09	8.86%	synonymous_variant
	chrA03__20116074	Exon	3.92E-09	9.09%	synonymous_variant
	chrA03__20116086	Exon	0.08171	1.25%	synonymous_variant
	chrA03__20116098	Exon	2.29E-09	9.69%	synonymous_variant
	chrA03__20116122	Exon	0.28067	0.66%	synonymous_variant
	chrA03__20116145	Exon	1.03E-09	10.10%	synonymous_variant
	chrA03__20116259	Exon	8.17E-15	15.06%	missense_variant
	chrA03__20116290	Exon	0.03405	1.70%	synonymous_variant
	chrA03__20116299	Exon	6.37E-11	11.28%	missense_variant
	chrA03__20116362	Exon	4.19E-05	5.15%	synonymous_variant
	chrA03__20116421	Exon	1.58E-05	4.93%	synonymous_variant
	chrA03__20116424	Exon	2.74E-05	4.66%	synonymous_variant
	chrA03__20116442	Exon	2.08E-04	4.56%	synonymous_variant
	chrA03__20116460	Exon	0.00383	2.97%	synonymous_variant
	chrA03__20116477	Exon	0.07348	1.44%	synonymous_variant
	chrA03__20116478	Exon	0.07711	1.42%	synonymous_variant
	chrA03__20116483	Exon	0.00114	2.88%	synonymous_variant
	chrA03__20116960	Exon	0.01848	2.31%	synonymous_variant
	chrA03__20116991	Exon	0.1227	1.16%	synonymous_variant
	chrA03__20117110	Exon	0.99056	0.01%	synonymous_variant
	chrA03__20117641	Exon	6.73E-16	16.50%	missense_variant

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrA03__20117699	Exon	0.55387	0.30%	synonymous_variant
	chrA03__20117809	Exon	0.96394	0.02%	synonymous_variant
	chrA03__20117836	Exon	0.38357	0.47%	synonymous_variant
	chrA03__20117839	Exon	0.36633	0.49%	synonymous_variant
	chrA03__20117891	Exon	0.56805	0.28%	synonymous_variant
	chrA03__20117911	Exon	1.47E-07	7.46%	synonymous_variant
	chrA03__20117925	Exon	0.52941	0.32%	synonymous_variant
	chrA03__20117955	Exon	0.45348	0.40%	synonymous_variant
	chrA03__20117976	Exon	0.58026	0.28%	synonymous_variant
	chrA03__20118008	Exon	0.60159	0.25%	synonymous_variant
	chrA03__20118021	Exon	0.15943	0.91%	synonymous_variant
	chrA03__20118123	Exon	0.42008	0.43%	synonymous_variant
	chrA03__20118126	Exon	0.53206	0.31%	synonymous_variant
	chrA03__20118135	Exon	0.81642	0.10%	synonymous_variant
	chrA03__20118158	Exon	1.98E-12	12.18%	missense_variant
	chrA03__20118184	Exon	0.18234	0.83%	synonymous_variant
	chrA03__20118200	Exon	0.51167	0.33%	synonymous_variant
	chrA03__20118207	Exon	0.37412	0.48%	synonymous_variant
	chrA03__20118237	Exon	0.76082	0.13%	synonymous_variant
	chrA03__20118320	Exon	0.73743	0.15%	synonymous_variant
	chrA03__20118338	Exon	1.48E-20	18.48%	missense_variant
	chrA03__20118340	Exon	2.50E-13	13.02%	missense_variant
	chrA03__20118437	Promoter	0.77546	0.13%	/
	chrA03__20118470	Promoter	0.95494	0.02%	/
	chrA03__20118477	Promoter	0.82013	0.10%	/

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrA03__20118479	Promoter	0.82013	0.10%	/
	chrA03__20118493	Promoter	0.90993	0.05%	/
	chrA03__20118518	Promoter	0.33321	0.55%	/
	chrA03__20118532	Promoter	0.80332	0.11%	/
	chrA03__20118537	Promoter	0.35581	0.52%	/
	chrA03__20118574	Promoter	0.16445	0.91%	/
	chrA03__20118579	Promoter	0.26789	0.66%	/
	chrA03__20118581	Promoter	0.1006	1.14%	/
	chrA03__20118582	Promoter	0.27819	0.64%	/
	chrA03__20118601	Promoter	5.18E-08	8.07%	/
	chrA03__20118620	Promoter	0.36723	0.51%	/
	chrA03__20118669	Promoter	0.70481	0.18%	/
	chrA03__20118676	Promoter	0.71329	0.18%	/
	chrA03__20118678	Promoter	8.32E-17	16.74%	/
	chrA03__20118703	Promoter	0.78203	0.13%	/
	chrA03__20118727	Promoter	7.80E-08	8.52%	/
	chrA03__20118737	Promoter	0.33959	0.57%	/
	chrA03__20118755	Promoter	1.84E-07	7.89%	/
	chrA03__20118770	Promoter	0.20796	0.88%	/
	chrA03__20119381	Promoter	0.92158	0.05%	/
	chrA03__20119537	Promoter	0.67827	0.24%	/
BnaC02g42260D	chrC02__44933063	Exon	5.45E-16	16.98%	missense_variant
	chrC02__44933117	Exon	0.0348	1.78%	synonymous_variant
	chrC02__44933150	Exon	0.0953	1.26%	synonymous_variant
	chrC02__44933198	Exon	0.02328	1.99%	synonymous_variant

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrC02__44933210	Exon	0.0162	2.18%	synonymous_variant
	chrC02__44933216	Exon	0.01058	2.43%	synonymous_variant
BnaC07g31100D	chrC07__35312872	Promoter	1.41E-17	20.11%	/
	chrC07__35313045	Promoter	6.46E-17	18.77%	/
	chrC07__35313091	Promoter	9.49E-18	19.58%	/
	chrC07__35313102	Promoter	4.04E-17	18.89%	/
	chrC07__35313202	Promoter	0.08381	0.79%	/
	chrC07__35313220	Promoter	0.44062	0.15%	/
	chrC07__35313244	Promoter	0.73862	0.15%	/
	chrC07__35313278	Promoter	0.71639	0.03%	/
	chrC07__35313289	Promoter	0.37971	0.48%	/
	chrC07__35313316	Promoter	0.34946	0.52%	/
	chrC07__35313322	Promoter	0.61957	0.06%	/
	chrC07__35313346	Promoter	7.25E-04	3.87%	/
	chrC07__35313357	Promoter	0.04181	1.64%	/
	chrC07__35313370	Promoter	0.00848	2.49%	/
	chrC07__35313380	Promoter	1.51E-20	20.54%	/
	chrC07__35313382	Promoter	0.8077	0.02%	/
	chrC07__35313424	Promoter	2.40E-05	5.37%	/
	chrC07__35313427	Promoter	0.06598	1.39%	/
	chrC07__35313514	Promoter	0.65624	0.23%	/
	chrC07__35313553	Promoter	1.52E-17	18.27%	/
	chrC07__35313574	Promoter	2.36E-17	19.37%	/

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrC07__35313577	Promoter	0.28848	0.69%	/
	chrC07__35313608	Promoter	0.64909	0.25%	/
	chrC07__35313661	Promoter	0.42571	0.52%	/
	chrC07__35313672	Promoter	0.67901	0.23%	/
	chrC07__35313681	Promoter	0.57681	0.10%	/
	chrC07__35313820	Promoter	6.82E-16	17.85%	/
	chrC07__35314061	Promoter	5.52E-27	27.53%	/
	chrC07__35314173	Promoter	1.24E-21	22.61%	/
	chrC07__35314338	Promoter	2.68E-21	22.31%	/
	chrC07__35314992	Promoter	9.26E-19	20.90%	/
	chrC07__35315124	Promoter	8.36E-18	19.58%	/
	chrC07__35315153	Promoter	5.56E-21	21.45%	/
	chrC07__35315177	Promoter	4.33E-05	5.79%	/
	chrC07__35315209	Promoter	1.44E-04	5.05%	/
	chrC07__35315539	Promoter	6.24E-24	23.89%	/
	chrC07__35315546	Promoter	2.12E-15	15.95%	/
	chrC07__35315557	Promoter	0.1243	1.10%	/
	chrC07__35315595	Promoter	9.76E-22	22.81%	/
	chrC07__35315612	Promoter	0.01198	2.41%	/
	chrC07__35315623	Promoter	0.01248	2.50%	/
	chrC07__35315670	Promoter	5.20E-23	24.13%	/

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrC07__35315710	Promoter	0.01191	1.84%	/
	chrC07__35315753	Promoter	4.38E-17	18.24%	/
	chrC07__35315763	Promoter	6.62E-16	17.12%	/
	chrC07__35315774	Promoter	0.47389	0.15%	/
	chrC07__35315794	Promoter	0.73232	0.03%	/
	chrC07__35316286	Promoter	0.00119	3.71%	/
	chrC07__35316597	Promoter	1.37E-19	21.14%	/
	chrC07__35316662	Promoter	1.02E-21	23.71%	/
	chrC07__35316777	Promoter	3.30E-22	24.84%	/
	chrC07__35316781	Promoter	1.26E-21	24.25%	/
	chrC07__35316827	Promoter	2.25E-19	21.37%	/
	chrC07__35316984	Exon	9.71E-21	21.17%	missense_variant
	chrC07__35317018	Exon	1.02E-20	22.09%	synonymous_variant
	chrC07__35317045	Exon	4.12E-08	9.06%	synonymous_variant
	chrC07__35317048	Exon	4.12E-08	9.06%	synonymous_variant
	chrC07__35317069	Exon	0.02142	2.23%	synonymous_variant
	chrC07__35317141	Exon	0.01185	2.43%	synonymous_variant
	chrC07__35317147	Exon	1.43E-17	18.83%	synonymous_variant
	chrC07__35317150	Exon	1.03E-17	18.98%	synonymous_variant
	chrC07__35317204	Exon	1.54E-17	19.49%	synonymous_variant

Candidate gene	SNP	SNP location	*p* value	PVE	Variant type
	chrC07__35317389	3'UTR	1.15E-17	20.76%	/

For *BnaA03g40240D*, candidate-gene association analysis detected several significant SNPs associated with SGC, including one promoter variant and multiple exonic variants (Fig. 5A; Table 1). Haplotype analysis based on seven representative SNPs classified the accessions into four major haplotypes (Fig. 5B). *BnaA03g40240D*-Hap1, represented by 23 accessions, showed the highest mean SGC of 102.00 μmol/g. By contrast, Hap2, Hap3, and Hap4 displayed markedly lower mean SGC values of 39.03, 37.51, and 30.85 μmol/g, respectively. Statistical comparison indicated that Hap2–Hap4 were significantly different from Hap1 and were associated with reduced SGC accumulation (Fig. 5C). These results suggest that *BnaA03g40240D*-Hap2, Hap3, and Hap4 represent favorable low-SGC haplotypes.

For *BnaC02g42260D*, association analysis identified one highly significant SNP, chrC02_44933063, within the candidate region (Fig. 6A; Table 1). This SNP showed a strong association with SGC, and accessions carrying the C allele exhibited significantly lower SGC than those carrying theT allele ($P = 5.7 \times 10^{-16}$; Fig. 6B). Thus, the C allele of chrC02_44933063 was considered a favorable allele and may serve as a useful marker for marker-assisted selection of low-SGC rapeseed lines.

For *BnaC07g31100D*, a total of fifteen SNPs were significantly associated with SGC, most of which were located in the promoter region, suggesting that regulatory variation may contribute to phenotypic differences at this locus (Fig. 7A; Table 1). Haplotype analysis using these SNPs divided the accessions into four major haplotypes (Fig. 7B). *BnaC07g31100D*-Hap1, represented by 71 accessions, had a high mean SGC of 109.97 μmol/g. In contrast, Hap2, Hap3, and Hap4 had much lower mean SGC values of 38.54, 39.54, and 41.25 μmol/g, respectively. Phenotypic comparisons showed that these three haplotypes were significantly associated with reduced SGC compared with Hap1 (Fig. 7C). Therefore, *BnaC07g31100D*-Hap2, Hap3, and Hap4 were identified as favorable haplotypes for low-SGC improvement.

Collectively, candidate-gene association and haplotype analyses demonstrated that natural allelic variation in these four candidate genes contributes substantially to SGC variation in the *B. napus* natural population. The favorable haplotypes or alleles identified here, including *BnaA09g05670D*-Hap1, *BnaA03g40240D*-Hap2/Hap3/Hap4, the C allele of chrC02_44933063 in *BnaC02g42260D*, and *BnaC07g31100D*-Hap2/Hap3/Hap4, provide valuable molecular targets for breeding low-glucosinolate rapeseed varieties.

3.5 Phenotypic characterization and breeding potential of integrated low-SGC haplotypes

To evaluate the breeding value of the favorable alleles identified above, we integrated the low-SGC haplotypes from the four candidate genes and compared their effects on seed quality, agronomic

performance, and root architecture. Accessions carrying the favorable low-SGC haplotype combinations were classified as the LSGC-hap group, whereas accessions carrying unfavorable high-SGC haplotype combinations were classified as the HSGC-hap group. In total, 38 accessions were assigned to the LSGC-hap group and 57 accessions to the HSGC-hap group.

As expected, the LSGC-hap group showed a markedly reduced seed glucosinolate content compared with the HSGC-hap group. The mean SGC decreased from 108.96 μmol/g in HSGC-hap accessions to 32.42 μmol/g in LSGC-hap accessions, representing a reduction of approximately 70.2% (Table 2; Fig. 8). This result confirms that pyramiding the favorable haplotypes identified from *BnaA09g05670D*, *BnaA03g40240D*, *BnaC02g42260D*, and *BnaC07g31100D* can substantially reduce SGC in *B. napus*.

Table 2
Phenotypic statistics of SGC, seed yield and yield related traits, root related traits, and seed oil content between *Brassic napus* varieties carrying "*LSGCHap*" and "*HSGCHap*"

	Trait	Mean	Median	Mad	Min	Max	Range
HSGCHap	SGC	108.96	125.99	31.45	14.20	187.07	172.87
	SOC	38.55	38.05	2.76	33.90	49.04	15.14
	C18:1	34.26	26.14	16.04	7.86	70.18	62.32
	C18:2	15.68	15.02	2.22	11.54	22.29	10.75
	C18:3	9.18	9.00	0.81	7.02	12.34	5.32
	C22:1	20.87	26.80	14.11	0.00	44.25	44.25
	SY	17.49	17.42	5.26	8.45	28.45	20.00
	SW	3.79	3.89	0.52	2.82	4.72	1.91
	PH	167.72	166.67	16.56	138.67	201.83	63.17
	BN	6.92	6.83	1.24	4.67	10.00	5.33
	PRL	12.98	13.30	2.70	7.58	17.42	9.84
	LRL	34.56	33.14	9.19	19.38	50.00	30.62
	TRL	46.62	46.03	13.40	25.53	66.32	40.79
	LRD	1.29	1.29	0.14	0.98	1.72	0.74
	LRN	16.53	16.45	5.00	8.11	24.82	16.71
	MLRL	0.70	0.70	0.03	0.63	0.78	0.16
LSGCHap	SGC	32.42	28.46	12.16	6.99	97.35	90.36
	SOC	40.85	40.83	3.78	33.63	48.08	14.45
	C18:1	60.43	60.65	5.00	40.33	73.90	33.57
	C18:2	18.71	18.80	1.10	15.27	21.46	6.19
	C18:3	9.75	9.67	0.85	7.42	13.16	5.74
	C22:1	0.64	0.00	0.00	0.00	12.70	12.70
	SY	17.42	18.01	5.23	6.96	29.95	22.99

Note: SGC, seed glucosinolate content (mol/g), SOC, seed oil content (%), C18:1, oleic acid, C18:2, linoleic acid, C18:3, linolenic acid, C22:1, erucic acid, SY, seed yield (g/plant), SW, seed weight (g/1000 seeds), PH, plant height (cm), BN, branch number (No./plant), PRL, primary root length (cm), LRL, lateral root length (cm), TRL, total root length (cm), LRD, lateral root density (cm/cm^2), LRN, lateral root number (No./plant), MLRL, mean lateral root length (cm).

SW	3.87	3.87	0.74	2.81	4.75	1.94
PH	164.98	164.50	17.79	134.33	192.67	58.33
BN	6.97	6.83	1.24	4.33	10.50	6.17
PRL	12.91	12.84	2.62	8.78	18.18	9.40
LRL	33.24	32.36	10.29	15.51	52.54	37.03
TRL	45.99	45.05	11.88	26.59	70.97	44.38
LRD	1.28	1.30	0.16	1.01	1.52	0.51
LRN	16.52	17.13	4.39	9.04	23.62	14.58
MLRL	0.69	0.70	0.05	0.60	0.78	0.18

Note: SGC, seed glucosinolate content (mol/g), SOC, seed oil content (%), C18:1, oleic acid, C18:2, linoleic acid, C18:3, linolenic acid, C22:1, erucic acid, SY, seed yield (g/plant), SW, seed weight (g/1000 seeds), PH, plant height (cm), BN, branch number (No./plant), PRL, primary root length (cm), LRL, lateral root length (cm), TRL, total root length (cm), LRD, lateral root density (cm/cm^2), LRN, lateral root number (No./plant), MLRL, mean lateral root length (cm).

We further investigated whether the integrated low-SGC haplotypes were associated with changes in other seed quality traits. Compared with the HSGC-hap group, the LSGC-hap group exhibited significantly higher seed oil content, increasing from 38.55% to 40.85% ($P < 0.01$; Table 2; Fig. 8). In addition, the fatty acid profile was markedly improved in the LSGC-hap group. Oleic acid content was significantly increased in LSGC-hap accessions compared with HSGC-hap accessions, whereas erucic acid content was dramatically reduced from 20.87% to 0.64% ($P < 0.0001$; Fig. 8). Linoleic acid and linolenic acid contents were also significantly higher in the LSGC-hap group than in the HSGC-hap group (Fig. 8). These results indicate that the favorable low-SGC haplotype combination is not only associated with reduced glucosinolate accumulation, but also with improved oil content and fatty acid composition.

To determine whether the reduction in SGC was accompanied by undesirable effects on agronomic performance, we compared several yield-related traits between the two haplotype groups. No significant differences were observed for seed yield, seed weight, plant height, or branch number between LSGC-hap and HSGC-hap accessions (Table 2; Fig. 8). For example, seed yield was comparable between the two groups, with mean values of 17.42 and 17.49 g per plant in the LSGC-hap and HSGC-hap groups, respectively. Similarly, plant height and branch number showed no significant differences. These findings suggest that the integrated low-SGC haplotypes reduce seed glucosinolate content without causing detectable penalties in major agronomic traits.

Because root architecture can influence nutrient acquisition and plant performance, we also compared root-related traits between the two haplotype groups. Lateral root density, lateral root length, lateral root number, mean lateral root length, primary root length, and total root length were all statistically

indistinguishable between LSGC-hap and HSGC-hap accessions (Table 4; Fig. 8). This indicates that the favorable low-SGC haplotype combination does not adversely affect root development.

Collectively, these results demonstrate that the integrated LSGC-hap genotype effectively reduces seed glucosinolate content while maintaining agronomic performance and root architecture. Moreover, this haplotype combination is associated with improved seed oil content, increased oleic acid content, and substantially reduced erucic acid content. Therefore, the LSGC-hap combination represents a valuable genetic module for breeding low-glucosinolate, high-oil-quality rapeseed cultivars without compromising yield-related traits.

4. Discussion

Seed glucosinolate content is a key quality trait in *Brassica napus*, directly affecting the nutritional and feed value of rapeseed meal. In this study, we combined multi-environment phenotyping, genome-wide association analysis, BnIR-based expression–phenotype association, and haplotype dissection to elucidate the genetic basis of SGC variation in a natural population of *B. napus*. The association panel exhibited extensive natural variation in SGC across environments, together with strong inter-environment correlations and high broad-sense heritability. These results indicate that SGC is a highly heritable trait and that this population provides a suitable genetic resource for dissecting allelic variation controlling seed glucosinolate accumulation.

Through GWAS using high-density SNP markers, we identified 11 significant QTLs distributed across multiple chromosomes, including A01, A02, A03, A08, A09, C02, C03, C07, C08, and C09. The detection of multiple loci is consistent with the quantitative inheritance of SGC in oilseed rape and supports previous findings that glucosinolate accumulation is controlled by a complex genetic network involving biosynthesis, transport, degradation, and seed-specific allocation processes (Chao et al. 2022; Liu et al. 2020). Several strong and stable association signals were repeatedly detected across different environments and in the BLUP dataset, suggesting that these loci have robust effects on SGC variation (Fig. 2). Compared with earlier studies, the loci identified here provide both confirmation of known genomic regions and additional candidate regions that may contribute to natural SGC diversity. These differences may be attributable to the broader genetic diversity of the association panel, environmental variation, marker density, and the statistical power provided by the high-density SNP dataset and the FaST-LMM model.

To prioritize candidate genes within stable GWAS loci, we further used the gene expression module of the BnIR database to evaluate the relationship between seed transcript abundance and SGC across the population. This strategy allowed us to narrow the candidate list from positional genes within GWAS intervals to genes whose expression levels were significantly associated with phenotypic variation. Four genes, *BnaA03g40240D*, *BnaA09g05670D*, *BnaC02g42260D*, and *BnaC07g31100D*, were prioritized as high-confidence candidates. Notably, *BnaC02g42260D* was annotated as *BnaC02.GTR2a*, a glucosinolate transporter gene involved in glucosinolate translocation and seed accumulation. Previous studies have

shown that GTR transporters participate in the allocation of glucosinolates among source and sink tissues and play important roles in determining seed glucosinolate content (Tan et al. 2022). The recovery of this known SGC-related gene supports the reliability of our GWAS and BnIR-based prioritization strategy. In contrast, *BnaA03g40240D*, *BnaA09g05670D*, and *BnaC07g31100D* have not been previously reported as major regulators of SGC, suggesting that they may represent novel loci contributing to seed glucosinolate variation in *B. napus* (Fig. 2).

The expression–phenotype correlations of these candidate genes also revealed potential stage-specific regulatory patterns. *BnaA09g05670D*, *BnaC02g42260D*, and *BnaC07g31100D* showed significant positive correlations with SGC at 20 DAF, whereas BnaA03g40240D was negatively correlated with SGC at 40 DAF (Fig. 3). These results suggest that different candidate genes may act at distinct seed developmental stages. Early seed development may be a critical period for glucosinolate import, biosynthesis, or allocation, while later developmental stages may involve redistribution, storage, or metabolic adjustment. The strong positive correlation between BnaC02g42260D/BnaC02.GTR2a expression and SGC is particularly consistent with its predicted role as a glucosinolate transporter. For the other three genes, further functional validation will be required to determine whether they directly regulate glucosinolate metabolism or influence SGC indirectly through related developmental or metabolic pathways.

Candidate-gene association and haplotype analyses provided further evidence that natural allelic variation in these four genes contributes to SGC diversity. For *BnaA09g05670D*, the Hap1 haplotype was associated with substantially lower SGC than the other haplotypes and was therefore identified as a favorable low-SGC haplotype (Fig. 4). In *BnaA03g40240D*, Hap2, Hap3, and Hap4 were associated with significantly reduced SGC compared with Hap1 (Fig. 5). For *BnaC02g42260D*, a single significant SNP, chrC02_44933063, showed a strong association with SGC, and the C allele was associated with markedly reduced SGC (Fig. 6). In *BnaC07g31100D*, Hap2, Hap3, and Hap4 were favorable haplotypes associated with low SGC, whereas Hap1 conferred high SGC. These results demonstrate that favorable alleles at multiple loci can be used to reduce SGC and that both coding and regulatory variation may contribute to phenotypic differences (Fig. 7). In particular, the large number of significant promoter variants detected in BnaC07g31100D suggests that expression-level regulation may be an important mechanism underlying SGC variation at this locus.

The integration of favorable low-SGC haplotypes further demonstrated the practical value of these loci for molecular breeding. Accessions carrying the integrated low-SGC haplotype combination showed a marked reduction in SGC compared with accessions carrying high-SGC haplotypes. This result supports the additive or cumulative contribution of multiple favorable loci to SGC reduction and suggests that pyramiding favorable haplotypes can achieve predictable improvement of this quantitative trait. Such a haplotype-based strategy is particularly suitable for SGC improvement because glucosinolate content is controlled by multiple genes with moderate or large effects rather than by a single locus. The favorable haplotypes and diagnostic SNPs identified in this study can be converted into breeder-friendly markers, such as KASP markers, for high-throughput marker-assisted selection.

An important finding of this study is that the integrated low-SGC haplotypes were not associated with unfavorable effects on major agronomic traits or root architecture. A common concern in rapeseed quality breeding is that reducing anti-nutritional compounds may be accompanied by yield penalties or compromised plant development. However, in our population, the LSGC-hap group showed no significant differences from the HSGC-hap group in seed yield, seed weight, plant height, branch number, lateral root traits, primary root length, or total root length. These results indicate that the identified haplotype combination can reduce SGC without detectable negative effects on plant performance. This decoupling of SGC reduction from agronomic penalties increases the breeding value of the favorable haplotypes.

In addition to reducing SGC, the LSGC-hap group exhibited improved seed quality traits, including higher seed oil content, increased oleic, linoleic, and linolenic acid contents, and strongly reduced erucic acid content. These results suggest that accessions carrying integrated low-SGC haplotypes may also possess favorable genetic backgrounds for oil quality improvement. The substantial reduction in erucic acid is particularly valuable because low erucic acid and low glucosinolate content are the defining characteristics of double-low rapeseed. However, because these quality traits may also be influenced by breeding history, population structure, and linkage with known oil-quality loci, the relationship between the low-SGC haplotype combination and fatty acid composition should be further validated in segregating populations or near-isogenic backgrounds. Nevertheless, the simultaneous improvement of SGC, oil content, and fatty acid composition observed in the present panel highlights the potential of haplotype-based selection to combine multiple quality traits in rapeseed breeding.

The breeding application of haplotype-directed selection has been demonstrated in several *B. napus* improvement programs. For example, KASP markers targeting *BnFAD2* alleles have been developed for high-throughput selection of elevated oleic acid content in segregating populations (Fu et al. 2021), while allele-specific CAPS and SNaPshot markers have been used to develop high-oleic and low-linolenic inbred lines with favorable agronomic performance (Spasibionek et al. 2020). Similarly, KASP markers have accelerated the introgression of clubroot resistance into elite restorer lines while maintaining yield potential (Guo et al. 2022). These examples demonstrate the feasibility of converting key haplotypes or functional variants into practical molecular markers. In the same way, the favorable alleles identified in *BnaA09g05670D*, *BnaA03g40240D*, *BnaC02g42260D*, and *BnaC07g31100D* can be developed into marker systems for pyramiding low-SGC alleles in breeding populations.

Despite these advances, several questions remain to be addressed. First, although expression–phenotype association and candidate-gene association analyses provide strong evidence for these genes, functional validation is still required. Gene editing, overexpression, RNA interference, or transgenic complementation could be used to confirm their causal roles in regulating SGC. Second, the biological functions of *BnaA03g40240D*, *BnaA09g05670D*, and *BnaC07g31100D* need to be characterized in detail to determine whether they participate directly in glucosinolate metabolism, transport, transcriptional regulation, or seed developmental processes. Third, the favorable haplotypes should be tested in different genetic backgrounds and multi-location breeding trials to evaluate their stability and

breeding performance under diverse environments. Such validation will be essential before large-scale deployment in commercial rapeseed breeding.

In conclusion, this study elucidated the genetic architecture of SGC in *B. napus* by integrating multi-environment GWAS, BnIR-based expression–phenotype association, candidate-gene association, and haplotype analysis. We identified 11 significant QTLs and prioritized four candidate genes, including the known glucosinolate transporter *BnaC02.GTR2a* and three previously unreported candidate loci. Haplotype dissection revealed multiple favorable alleles that significantly reduce SGC, and pyramiding these favorable haplotypes produced a low-SGC genetic configuration without compromising yield-related or root developmental traits. Moreover, the integrated low-SGC haplotypes were associated with improved seed oil content and fatty acid composition. These findings provide valuable genetic resources and practical molecular targets for breeding low-glucosinolate, high-quality rapeseed cultivars.

Declarations

Declaration of interests

The authors declare no competing interests.

Author Contribution

Huafang Wan and Haijiang Liu designed the research, reviewed the writing, and drafted the manuscript. Kaijie Ye, Liang Ran and Qiuting Wei participated in the experiments. Nengwen Yin, Huiyan Zhao, Ti Zhang, Taocui Huang, Hai Du and Cunmin Qu participated in manuscript revision and data interpretation.

Acknowledgments

National Key R&D Program of China (2023YFF1000700), National Natural Science Foundation of China (32272150, 32372189), Chongqing Technology Innovation and Application Development Program (CSTB2023TIAD-KPX0038), Natural Science Foundation of Chongqing (CSTB2023NSCQ-MSX0744), the earmarked fund for CARS-12.

References

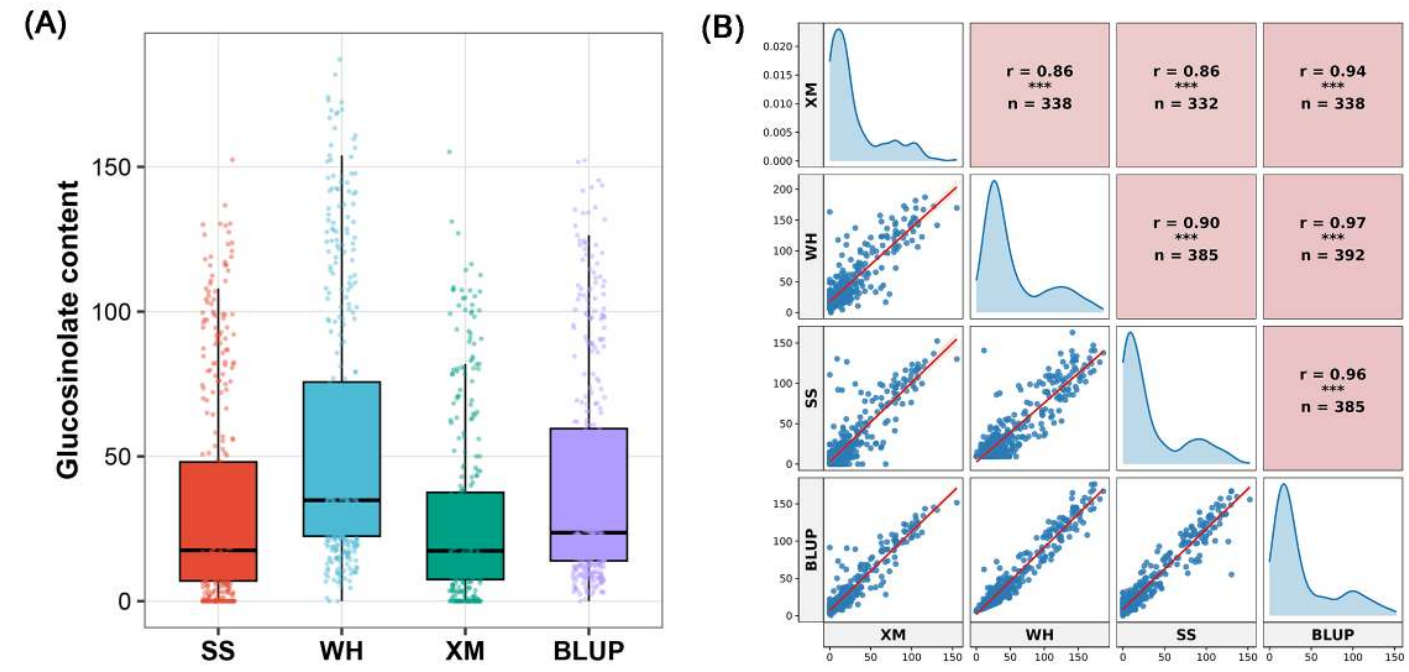
1. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265
2. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

3. Chao H, Li H, Yan S, Zhao W, Chen K, Wang H, Raboanatahiry N, Huang J, Li M (2022) Further insight into decreases in seed glucosinolate content based on QTL mapping and RNA-seq in *Brassica napus* L. Theor Appl Genet 135:2969–2991
4. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158
5. Fu Y, Mason AS, Zhang Y, Yu H (2021) Identification and Development of KASP Markers for Novel Mutant *BnFAD2* Alleles Associated With Elevated Oleic Acid in *Brassica napus*. Front Plant Sci 12:715633
6. Guo Y, Li B, Li M, Zhu H, Yang Q, Liu X, Qu L, Fan L, Wang T (2022) Efficient marker-assisted breeding for clubroot resistance in elite Pol-CMS rapeseed varieties by updating the *PbBa8.1* locus. Mol Breed 42:41
7. Hassan M (2014) Effectivenss of Gamma Rays to Induce Genetic Variability to Improve Some Agronomic Traits of Canola (*Brassica napus* L.). Asian Journal of Crop Science 6:123–132.
8. Hu D, Jing J, Snowdon RJ, Mason AS, Shen J, Meng J, Zou J (2021) Exploring the gene pool of *Brassica napus* by genomics-based approaches. Plant Biotechnol J 19:1693–1712
9. Liu H, Li X, Zhang Q, Yuan P, Liu L, King GJ, Ding G, Wang S, Cai H, Wang C, Xu F, Shi L (2021) Integrating a genome-wide association study with transcriptomic data to predict candidate genes and favourable haplotypes influencing *Brassica napus* seed phytate. DNA Res 28
10. Liu H, Pan Y, Cui R, Hammond JP, White PJ, Zhang Y, Zou M, Ding G, Wang S, Cai H, Xu F, Shi L (2023) Integrating genome-wide association studies with selective sweep reveals genetic loci associated with tolerance to low phosphate availability in *Brassica napus*. Mol Breed 43:53
11. Liu H, Wang J, Zhang B, Yang X, Yuan P, Ding G, Wang S, Cai H, Wang C, Xu F, Shi L (2022a) Genome-wide association study identifies new loci for 1000-seed weight in *Brassica napus*. Euphytica 218:142
12. Liu H, Zou M, Zhang B, Yang X, Yuan P, Ding G, Xu F, Shi L (2022b) Genome-wide association study identifies candidate genes and favorable haplotypes for seed yield in *Brassica napus*. Mol Breed 42:61
13. Liu HJ, Yuan YH, Ye KJ, Li RS, Ran H, Tang YS, Lu K, Yin NW, Zhao HY, Li J, Huang TC, Shi L, Du H, Qu CM (2025) Dissection of the genetic architecture of seed erucic acid content and consequences for breeding in *Brassica napus* L. Industrial Crops and Products 226
14. Liu Y, Zhou X, Yan M, Wang P, Wang H, Xin Q, Yang L, Hong D, Yang G (2020) Fine mapping and candidate gene analysis of a seed glucosinolate content QTL, qGSL-C2, in rapeseed (*Brassica napus* L.). Theor Appl Genet 133:479–490
15. Siddiqui M, Khan I, Khatri A (2009) Induced quantitative variability by gamma rays and ethylmethane sulphonate alone and in combination in rapeseed (*Brassica napus* L.). Pakistan Journal of Botany 41
16. Spasibionek S, Mikołajczyk K, Ćwiek-Kupczyńska H, Piętka T, Krótka K, Matuszczak M, Nowakowska J, Michalski K, Bartkowiak-Broda I (2020) Marker assisted selection of new high oleic and low

linolenic winter oilseed rape (*Brassica napus* L.) inbred lines revealing good agricultural value. Plos One 15:e0233959

17. Tan Z, Xie Z, Dai L, Zhang Y, Zhao H, Tang S, Wan L, Yao X, Guo L, Hong D (2022) Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in *Brassica napus*. Plant Biotechnol J 20:211–225
18. Tang S, Zhao H, Lu SP, Yu LQ, Zhang GF, Zhang YT, Yang QY, Zhou YM, Wang XM, Ma W, Xie WB, Guo L (2021) Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. Molecular Plant 14:470–487
19. Voon PT, Ng CM, Ng YT, Wong YJ, Yap SY, Leong SL, Yong XS, Lee SWH (2024) Health Effects of Various Edible Vegetable Oil: An Umbrella Review. Adv Nutr 15:100276
20. Wang X, Chen Y, Thomas CL, Ding G, Xu P, Shi D, Grandke F, Jin K, Cai H, Xu F, Yi B, Broadley MR, Shi L (2017) Genetic variants associated with the root system architecture of oilseed rape (*Brassica napus* L.) under contrasting phosphate supply. DNA Res 24:407–417
21. Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide Association Studies in Maize: Praise and Stargaze. Mol Plant 10:359–374
22. Zhao C, Xie M, Liang L, Yang L, Han H, Qin X, Zhao J, Hou Y, Dai W, Du C, Xiang Y, Liu S, Huang X (2022) Genome-Wide Association Analysis Combined With Quantitative Trait Loci Mapping and Dynamic Transcriptome Unveil the Genetic Control of Seed Oil Content in *Brassica napus* L. Front Plant Sci 13:929197

Figures



(C)

Environment	Mean	Min	Max	Range	Skew	Kurtosis	CV value	Heritability
SS	33.27	0.01	152.46	152.46	1.22	0.24	111.50%	94.73%
WH	54.68	0.01	187.07	187.07	1.15	0.07	83.43%	
XM	29.05	0.01	155.18	155.18	1.43	1.17	109.77%	
BLUP	40.1	0.01	152.22	152.22	1.25	0.33	93.75%	

Figure 1

Phenotypic performance of seed glucosinolate content (SGC) in a *B. napus* natural population across multiple environments.(A) Phenotypic distribution of SGC in SS, WH, XM, and the BLUP values. (B) Pairwise correlation analysis of SGC across the three environments and BLUP values. The diagonal panels show the density distribution, the lower panels show scatter plots, and the upper panels show Pearson's correlation coefficients (r) and sample sizes (n). (C) Summary statistics of SGC in different environments and BLUP values, including mean, minimum, maximum, range, skewness, kurtosis, coefficient of variation, and broad-sense heritability. High correlations among environments and a high H^2 value (94.73%) indicate that SGC is a stable trait with strong genetic determination in the rapeseed natural population.

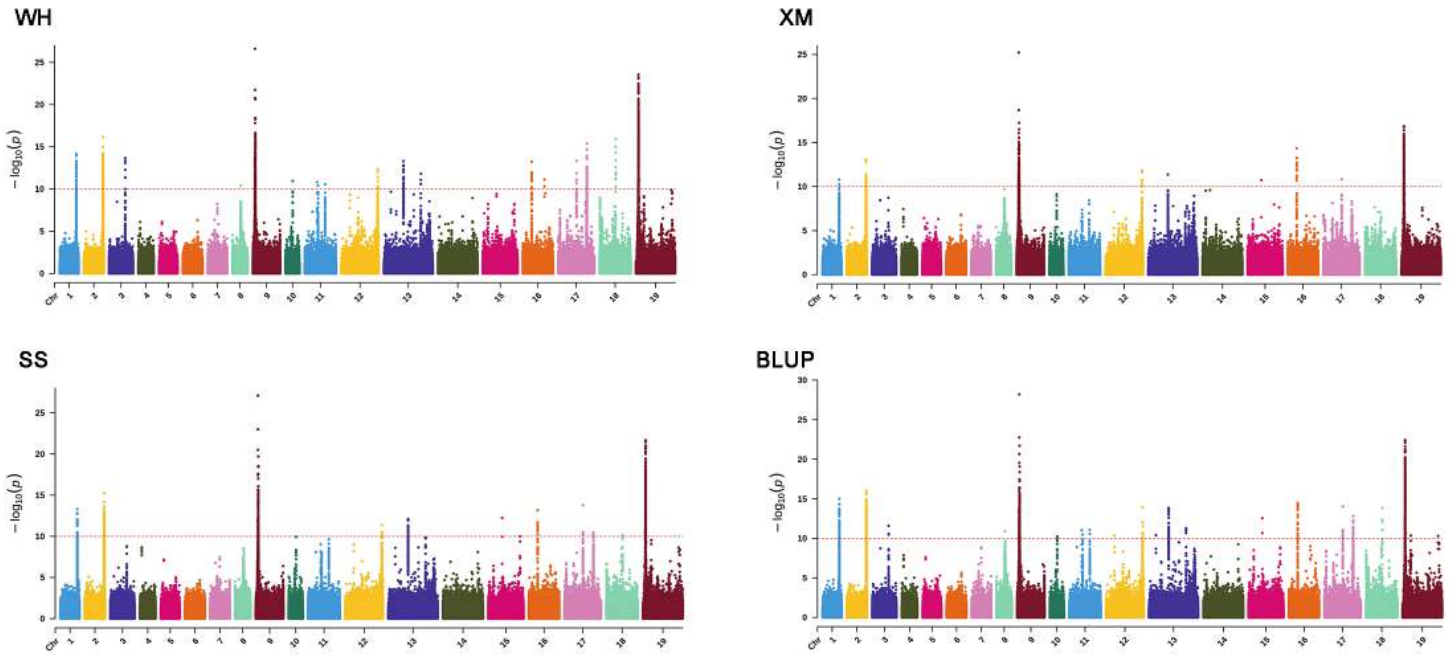


Figure 2

Genome-wide association analysis of seed glucosinolate content in *B. napus*. Manhattan plots showing GWAS signals for seed glucosinolate content (SGC) in a rapeseed natural population under three environments and BLUP values. The four panels correspond to WH, XM, SS, and BLUP, respectively. The x-axis shows the genomic positions of SNPs across the 19 chromosomes of *B. napus*, and the y-axis shows $-\log_{10}(P)$ values. The blue and red horizontal lines indicate the suggestive and genome-wide significance thresholds, respectively. Repeatedly detected association peaks across environments and BLUP values represent stable loci potentially involved in the genetic regulation of SGC.

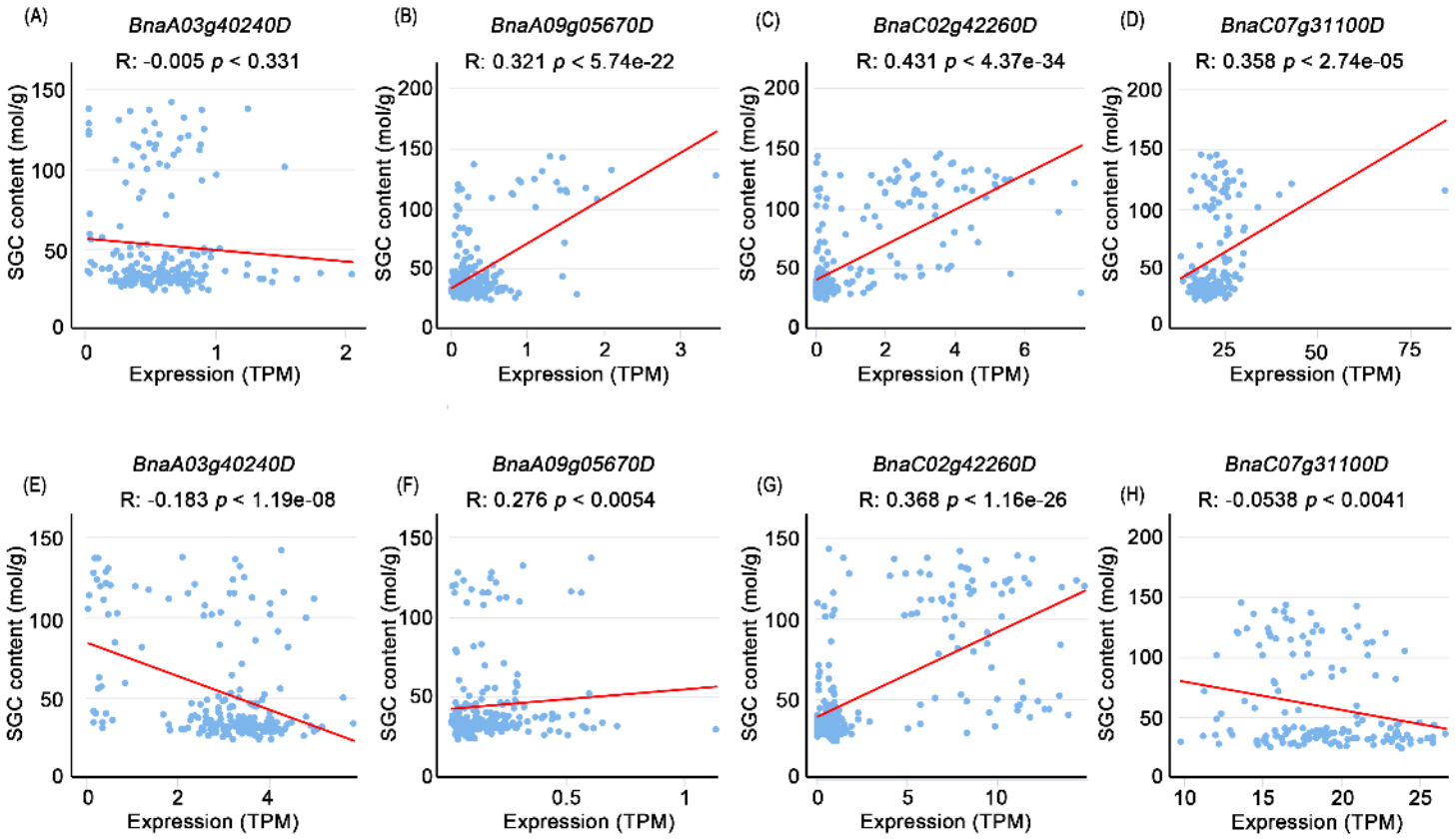
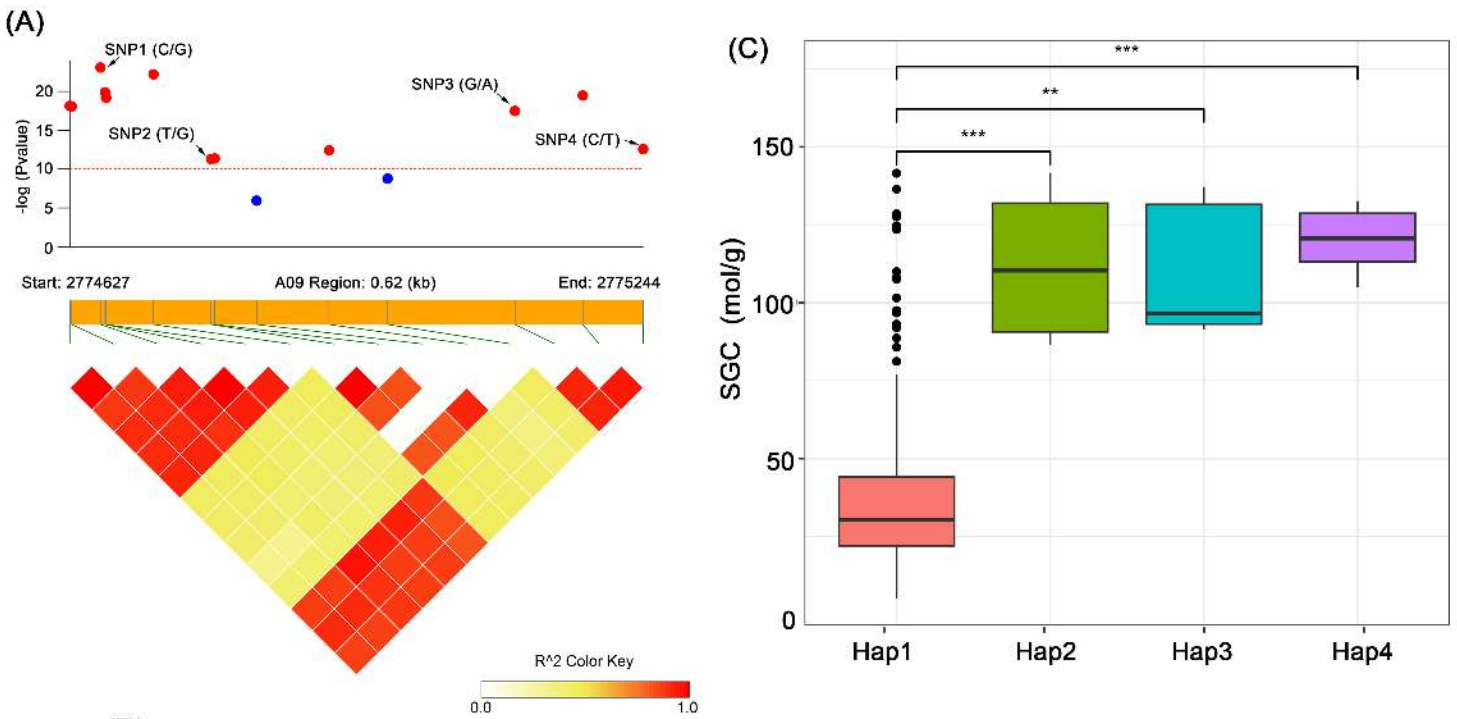


Figure 3

Expression–phenotype correlation analysis of candidate genes associated with seed glucosinolate content in *Brassica napus*. (A–H) Correlations between seed transcript abundance and seed glucosinolate content (SGC) for four candidate genes located within stable GWAS loci. Expression data were obtained from the BnIR database and are shown as transcripts per million (TPM). Scatter plots show the relationships between gene expression levels and SGC across the natural population at 20 days after flowering (DAF; A–D) and 40 DAF (E–H). The analyzed genes include *BnaA03g40240D* (A, E), *BnaA09g05670D*(B, F), *BnaC02g42260D* (C, G), and *BnaC07g31100D* (D, H). Pearson's correlation coefficients (R) and corresponding P values are indicated in each panel. Red lines represent linear regression fits.

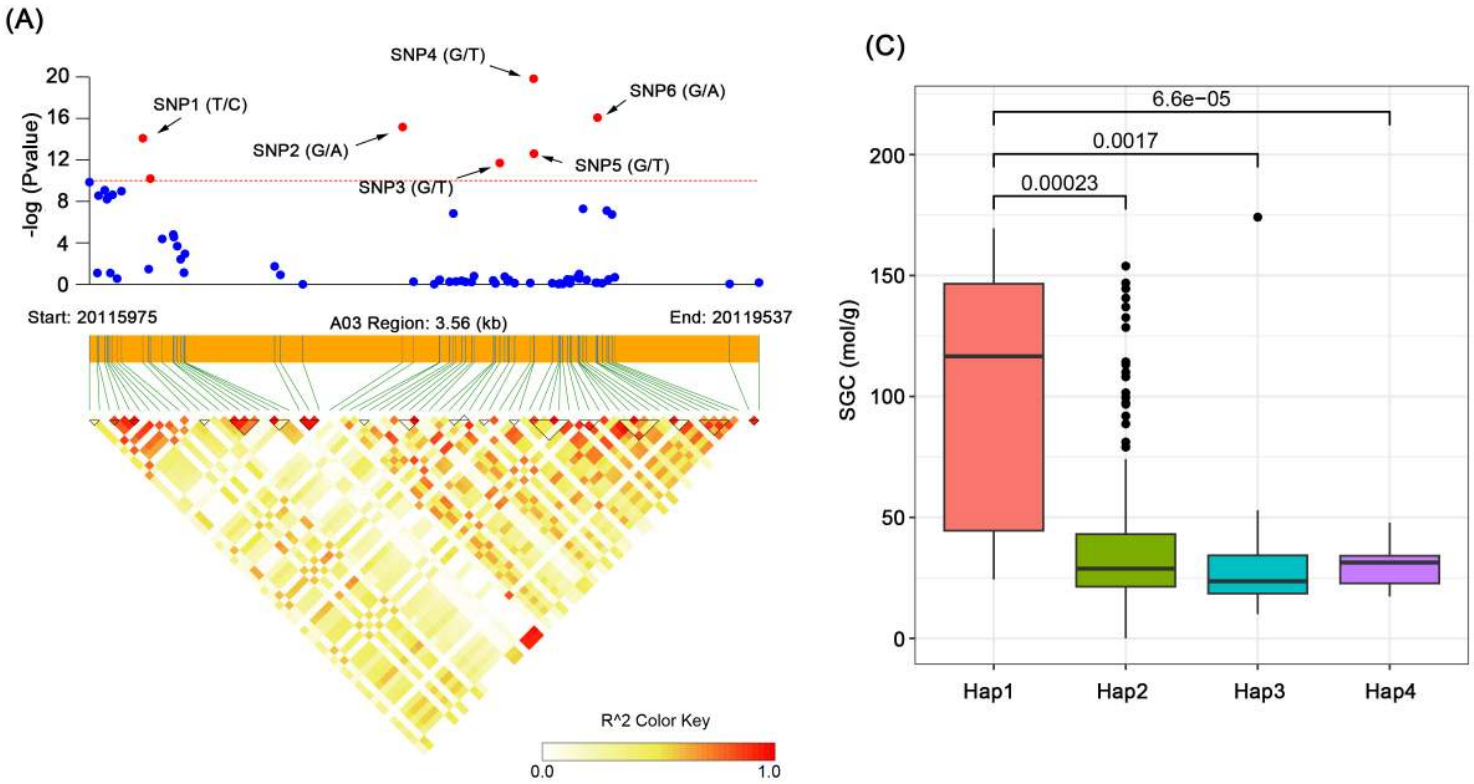


(B)

Haplotype	SNP1	SNP2	SNP3	SNP4	Number of lines	SGC content (mol/g)
Hap1	C	T	G	C	200	39.83
Hap2	G	G	A	T	14	118.58
Hap3	G	T	A	T	12	98.69
Hap4	G	T	A	C	11	120.04

Figure 4

Candidate-gene association and haplotype analysis of *BnaA09g05670D* for seed glucosinolate content. (A) Regional association analysis, gene structure, and linkage disequilibrium heatmap of the *BnaA09g05670D* locus. The upper panel shows the association signals of SNPs within the candidate region, with the y-axis representing $-\log_{10}(P)$. Significantly associated SNPs are highlighted and annotated. The red dashed line indicates the significance threshold. The middle panel shows the gene model, and the lower panel shows the pairwise LD pattern among SNPs. (B) Haplotype composition of *BnaA09g05670D* based on four lead SNPs. The number of accessions and mean SGC for each haplotype are shown. (C) Boxplots comparing SGC among the four haplotypes. Statistical significance was determined by multiple comparison tests, with P < 0.01 and *P < 0.001 indicated above the plots. Hap1 was associated with significantly lower SGC and was considered the favorable haplotype.



(B)

Haplotype	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	Number of lines	SGC content (mol/g)
Hap1	T	G	T	G	T	G	A	23	102
Hap2	C	A	G	T	G	T	G	213	39.03
Hap3	C	A	G	T	G	G	G	14	37.51
Hap4	C	G	G	T	G	T	G	13	30.85

Figure 5

Candidate-gene association and haplotype analysis of *BnaA03g40240D* for seed glucosinolate content. (A) Regional association plot, gene structure, and linkage disequilibrium heatmap of the *BnaA03g40240D*locus. The upper panel shows SNP–SGC association signals, with the y-axis representing $-\log_{10}(P)$. Significant SNPs are labeled, and the red dashed line denotes the significance threshold. The middle and lower panels show the gene model and LD structure, respectively. (B) Haplotype composition of *BnaA03g40240D*based on seven representative SNPs. The number of accessions and mean SGC for each haplotype are listed. (C) Phenotypic comparison of SGC among the four haplotypes. Hap1 showed significantly higher SGC, whereas Hap2, Hap3, and Hap4 were associated with reduced SGC and were considered favorable haplotypes. P values for significant comparisons are indicated above the boxplots.

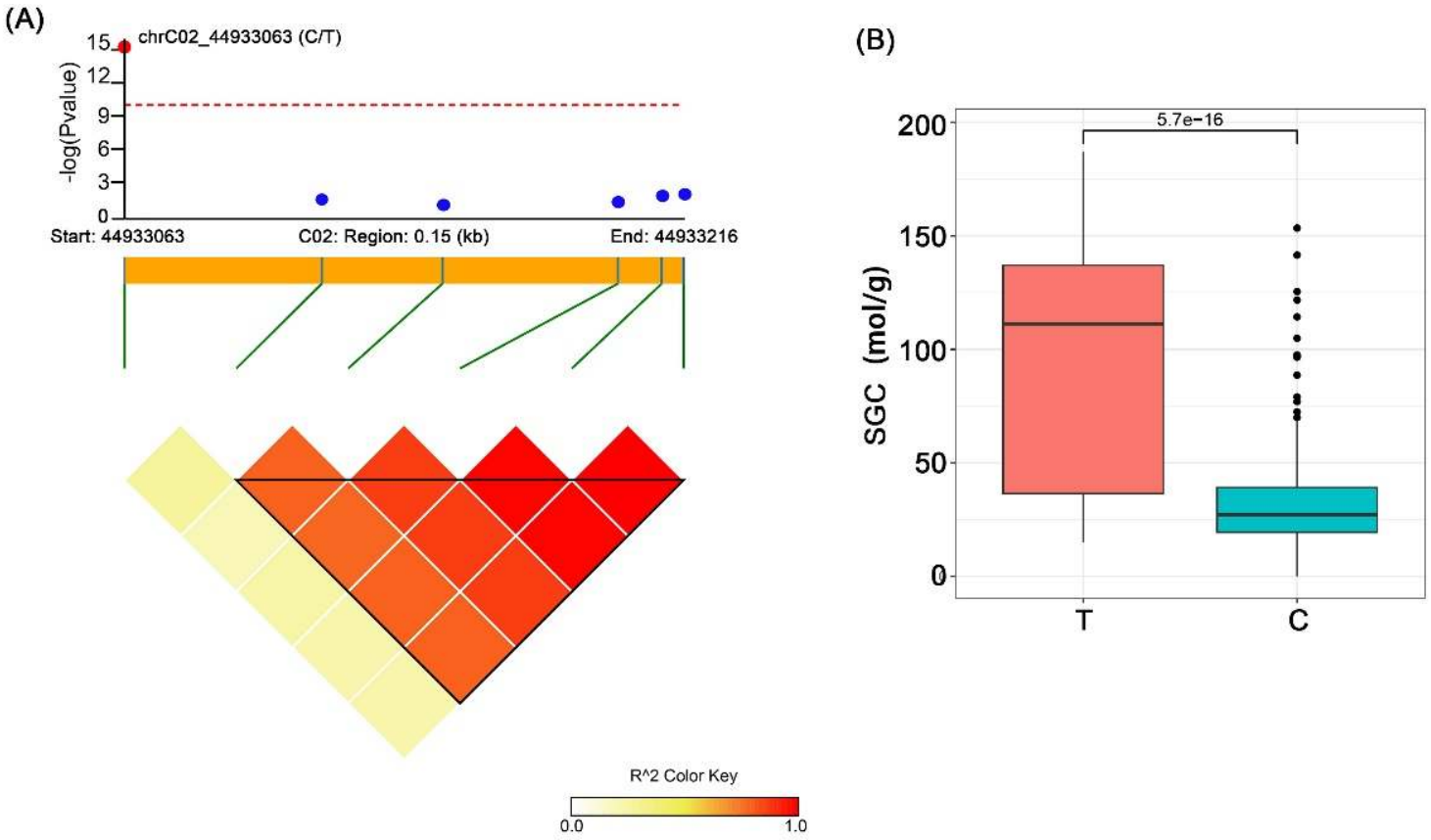
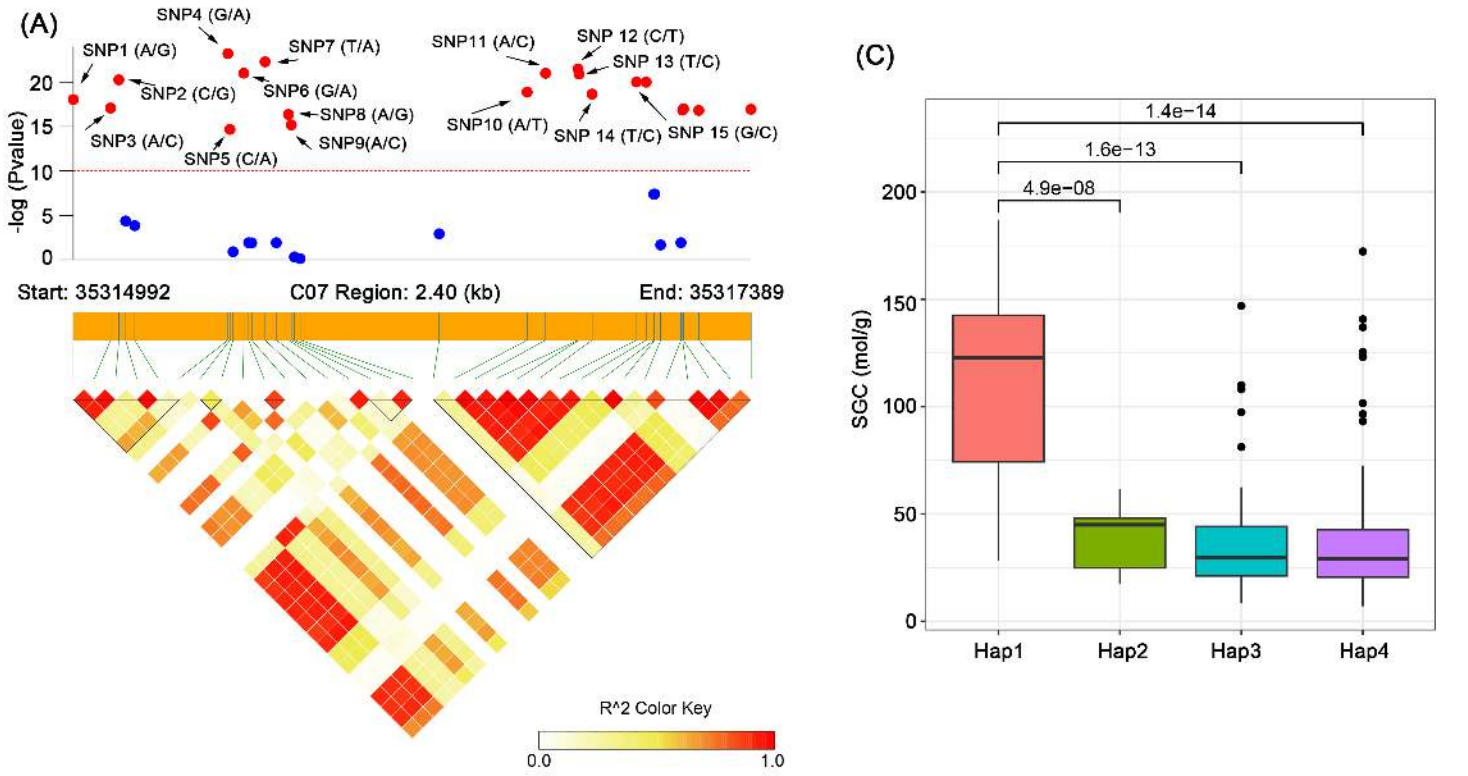


Figure 6

Candidate-gene association analysis of *BnaC02g42260D* identifies a favorable allele for low seed glucosinolate content. (A) Regional association analysis, gene structure, and linkage disequilibrium heatmap of the *BnaC02g42260D* locus. The upper panel shows SNP–SGC association signals, with the y-axis representing $-\log_{10}(P)$. The significant SNP chrC02_44933063 is annotated, and the red dashed line indicates the significance threshold. (B) Boxplots comparing SGC between accessions carrying the T and C alleles at chrC02_44933063. Accessions with the C allele showed significantly lower SGC than those with the T allele, indicating that the C allele represents a favorable allele for low-SGC breeding.



(B)

Haplotype	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	SNP11	SNP12	SNP13	SNP14	SNP15	Number of lines	SGC content (mol/g)
Hap1	A	C	A	G	C	G	T	A	A	A	A	C	T	T	G	71	109.97
Hap2	G	G	C	A	C	A	A	A	C	T	C	G	C	C	C	11	38.54
Hap3	G	G	C	A	C	A	A	G	C	T	C	G	C	C	C	47	39.54
Hap4	G	G	C	A	A	A	A	G	C	T	C	G	C	C	C	85	41.25

Figure 7

Candidate-gene association and haplotype analysis of *BnaC07g31100D* for seed glucosinolate content. (A) Regional association plot, gene structure, and linkage disequilibrium heatmap of the *BnaC07g31100D* locus. The upper panel shows association signals between SNPs and SGC, with the y-axis representing $-\log_{10}(P)$. Significant SNPs are annotated, and the red dashed line denotes the significance threshold. The middle panel shows the gene model, and the lower panel shows the pairwise LD pattern. (B) Haplotype composition of *BnaC07g31100D*based on fifteen significant SNPs. The number of accessions and mean SGC for each haplotype are provided. (C) Boxplots showing SGC differences among the four haplotypes. Hap1 was associated with high SGC, whereas Hap2, Hap3, and Hap4 showed significantly reduced SGC and were identified as favorable low-SGC haplotypes. P values for significant comparisons are indicated above the boxplots.

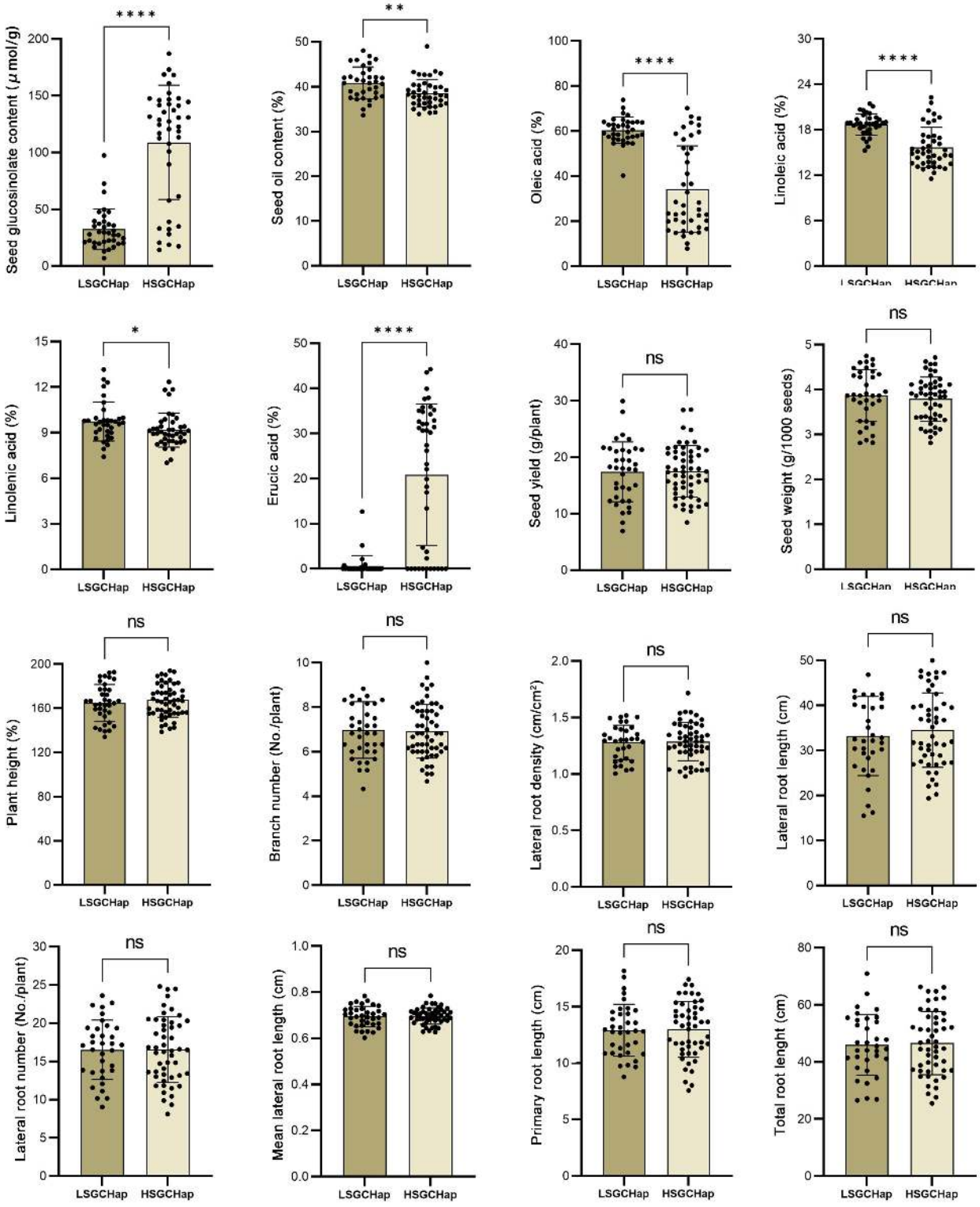


Figure 8

Phenotypic comparison of seed quality, agronomic performance, and root architecture between *B. napus* accessions carrying integrated low-SGC haplotypes (LSGC-hap) and high-SGC haplotypes (HSGC-hap). (A–H)Seed glucosinolate content (SGC), seed oil content (SOC), and fatty acid composition (oleic, linoleic, linolenic, and erucic acids). **(I–L)** Agronomic and yield-related traits: seed yield, seed weight, plant height, and branch number. **(M–P)** Root developmental metrics: lateral root density, lateral

root length, lateral root number, mean lateral root length, primary root length, and total root length. Group differences were evaluated for statistical significance using a t-test: $P < 0.05$ (*), $P < 0.01$ (**), $P < 0.0001$ (****); ns, not significant (*P* > 0.05). Sample sizes: LSGC-hap ($n = 38$), HSGC-hap ($n = 57$).