Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in *Brassica napus*

Zengdong Tan¹,†, Zhaoqi Xie¹,†, Lihong Dai¹, Yuting Zhang¹, Hu Zhao¹, Shan Tang¹, Lili Wan²*, Xuan Yao¹, Liang Guo¹,³,* and Dengfeng Hong¹,³,*
doi: 10.1111/pbi.13707

received 10 April 2021; revised 25 June 2021; accepted 7 September 2021.
*Correspondence (Tel +86 27 87286881; fax +86 27 87286881; email guoliang@mail.hzau.edu.cn (L.G.); Tel +86 27 87281713; fax +86 27 87384670; email dfhong@mail.hzau.edu.cn (D.H.))
†These authors contributed equally to this work.

Summary
A high content of seed glucosinolates and their degradation products imposes anti-nutritional effects on livestock; therefore, persistent efforts are made to reduce the seed GSL content to increase the commercial value of rapeseed meal. Here, we dissected the genetic structure of SGC by genome-wide association studies (GWAS) combined with transcriptome-wide association studies (TWAS). Fifteen reliable quantitative trait loci (QTLs) were identified to be associated with the reduced SGC in modern *B. napus* cultivars by GWAS. Analysis of the selection strength and haplotypes at these QTLs revealed that low SGC was predominantly generated by the co-selection of qGSL.A02.2, qGSL.C02.1, qGSL.A09.2, and qGSL.C09.1. Integration of the results from TWAS, comprehensive bioinformatics, and POCKET algorithm analyses indicated that *BnaC02.GTR2* (BnaC02g42260D) is a candidate gene underlying qGSL.C02.1. Using CRISPR/Cas9-derived Bna.gtr2s knockout mutants, we experimentally verified that both BnaC02.GTR2 and its three paralogs positively regulate seed GSL accumulation but negatively regulated vegetative tissue GSL contents. In addition, we observed smaller seeds with higher seed oil content in these Bna.gtr2 mutants. Furthermore, both RNA-seq and correlation analyses suggested that *Bna.GTR2*s might play a comprehensive role in seed development, such as amino acid accumulation, GSL synthesis, sugar assimilation, and oil accumulation. This study unravels the breeding selection history of low-SGC improvement and provides new insights into the molecular function of *Bna.GTR2*s in both seed GSL accumulation and seed development in *B. napus*.

Keywords: *Brassica napus*, glucosinolate, GTR2, GWAS, TWAS, co-expression, CRISPR, Cas9.

Introduction

Canola-type rapeseed (*Brassica napus*, AACC, *2n = 38*) is currently one of the main oil crops and a promising alternative to soybean source protein feed in the world (Chao et al., 2017; Gatrell et al., 2014; Wang et al., 2018). However, earlier unimproved rapeseed cultivars contain high seed glucosinolate (GSL) concentrations, which reduce the nutritional value and feed quality of the seed (Schmidt and Bancroft, 2011). ‘Bronowski,’ a feed landrace used in Poland in the late 1960s, provided the only germplasm source for breeding seed GSL content (SGC) in *B. napus* (Agnihotri et al., 2007). Subsequently, with the discovery of low-erucic cultivar ‘Liho’, the ‘double-low’ (low seed erucic acid and GSL content) *B. napus* varieties have been successfully developed and quickly popularized in major *B. napus* growing countries since the 1990s (Agnihotri et al., 2007; Burton et al., 2003a,b; Gupta and Pratap, 2007). However, glucosinolates (GSLs) are important defence compounds for resistance to herbivores and pathogens in plants (Wittstock et al., 2003). Reportedly, the efforts made to reduce the GSL contents in seeds have concurrently reduced the GSL contents in leaves and other tissues, resulting in an overall decline in disease and stress resistance in *B. napus* (Li et al., 1999; Mithen, 1992). Therefore, tissue-specific control of GSL accumulation is of great significance for the quality improvement of *B. napus* (Nour-Eldin et al., 2012).

The GSL biosynthesis pathway has been widely studied in the model plant *Arabidopsis thaliana* and has evolved to produce more than 130 GSL structures with multiple amino acid-derived side chains (Agerbirk and Olsen, 2012; Fahey et al., 2001). Based on different precursor amino acids for side chains, GSL can be classified as aliphatic (methionine, alanine, valine, leucine, or isoleucine), aromatic (tyrosine or phenylalanine), and indolic (tryptophan) GSL (Blazevic et al., 2020; Clarke, 2010). GSL biosynthesis can generally be divided into three stages: extension of the side chain, formation of the core structure, and modification. The *R*-side chain (Barco and Clay, 2019; Fahey et al., 2001). Many gene families including *BCAT*s (Lachler et al., 2015; Schuster et al., 2006), MAMs (Redovnikovic et al., 2012; Textor et al., 2007), CYP79 family (Hansen et al., 2001; Reintanz et al., 2001; Tantikanjana et al., 2004; Zang et al., 2008), CYP83 family...
GWAS or linkage association mapping methods have been used to identify relevant QTLs that regulate GSL biosynthesis and accumulation (Feng et al., 2012; Liu et al., 2020; Lu et al., 2014). Two to three major QTLs were found to control the SGC of several bi-parental mapping populations (Howell et al., 2003; Toroser et al., 1995; Zhao and Meng, 2003). Subsequently, Feng et al. (2012) identified 71 seed-specific metabolite QTLs for GSL compounds. Furthermore, 43 QTLs for SGC were mapped over seven environments in Germany and China in a doubled haploid population in B. napus (Fu et al., 2015). With the rapid development of the new biotechnology, especially GWAS and TWAS, a considerable progress in the exploration of seed GSL accumulation in B. napus has been made in recent years. Through associative transcriptomics, Lu et al. (2014) identified 10 SNPs closely linked to genes related to SGC. Furthermore, Wang et al. (2018) identified 49 loci related to SGC using high-throughput resequencing. Based on these findings, we hypothesize that combining genomic and transcriptomic approaches could unravel the genetic basis of SGC and provide insights into the application of transport engineering in B. napus breeding.

Therefore, this study aims to comprehensively dissect the genetic basis of SGC at both the genomic and transcriptomic levels in B. napus by GWAS, TWAS, and co-expression network analysis. We identified 15 QTLs, 1854 genes, and regulatory networks associated with SGC. In addition, BnaA02.GTR2 was found to be a key gene influencing seed GSL accumulation as well as multiple seed traits.

Results

GWAS deciphered QTLs associated with SGC in B. napus

To determine the QTLs associated with SGC of B. napus, SGC of 505 B. napus accessions (Tang et al., 2020) was examined in Wuhan (WH) for 3 years (2017−2019). The Pearson coefficients of determination ($R^2$) of seed GSL phenotype data estimated using the best linear unbiased prediction (BLUP) method were stable in WH 2017−2019 (0.87−0.91) with improved $R^2$ between the years (0.94−0.98; Figure S1).

We performed GWAS for SGC using a linear mixed model (LMM) in the genome-wide efficient mixed-model association (GEMMA) software (Buss et al., 2016; Figure S2). Significant association signals were detected using the threshold of 6.91 x 10^{-7} of BLUP value in the data of WH 2017−2019 (Dataset S1). Finally, 15 QTLs mainly located on chromosomes A02, A09, C02, and C09 were identified as stable QTLs, detected at least twice in 3 years (Datasets S2 and S1, Figure 1A). Of these QTLs, qGSL.A02, qGSL.A09, qGSL.C02, qGSL.C07, and qGSL.C09 were reported in previous studies (Feng et al., 2012; Harper et al., 2012; Liu et al., 2020). Approximately, in the 150 kb region covered by the 15 QTLs, we identified 669 genes as GWAS-associated candidate genes (Dataset S3).

TWAS unravelled the molecular basis of seed GSL accumulation

Twenty days after flowering (DAF) and 40 DAF are critical developmental stages for seed filling and embryo growth, with the accumulation of GSL in B. napus (Gijzen et al., 1989; Larden and Tribol-Blondel, 1994; Murphy and Cummins, 1989; Yu et al., 2010). To further explore the genetic basis of seed GSL accumulation, we performed TWAS using the seed transcriptome data of two populations comprising 309 accessions at 20 DAF and 274 accessions at 40 DAF (Tang et al., 2020). We found 851
and 1327 genes significantly associated with SGC at 20 and 40 DAF (false discovery rate, FDR < 0.05) at 20 DAF for SGC. The position of the genome is plotted on the X-axis and the log-transformed FDR values of association between gene expression and SGC are plotted on the Y-axis. The genes above the bold black line are positively associated with SGC, and the genes shown below are negatively associated. The dashed grey horizontal lines represent the significance level. (d) Venn diagram of overlapped genes significant in TWAS at 20 and 40 DAF. (e) GO enrichment analysis of significant genes identified by TWAS at 20 DAF. The count number represents the number of genes enriched in a process, and the colour represents the range of $P$-value.

Genes with the same or similar functions can usually be divided into modules based on their expression patterns, and analysis of the relationship between modules and traits could help discover key genes controlling the trait (Ma et al., 2013). Analysis of the relationship between modules and trait can help discover the key genes controlling the trait. We used weighted gene co-expression network analysis (WGCNA) to identify highly synergistic sets of genes for independent modules from GWAS and TWAS results (DiLeo et al., 2011; Langfelder and Horvath, 2008). It has been shown that scale-free networks can make gene sets more to identify biologically more meaningful gene sets (Hein et al., 2006). Therefore, we estimated the best soft-thresholding power ($\beta$) value for constructing the scale-free network at 20 and 40 DAF. The $\beta$ that allowed scale independence to exceed 0.85 for the first time was used as a threshold to classify the gene modules (Figure S4). At 20 and 40 DAF, 15 and 11 co-expression modules ($\beta_{20\text{DAF}} = 14$, $\beta_{40\text{DAF}} = 8$) were divided for further analysis.
At 20 DAF, three modules were significantly positively correlated with SGC (red: \( r = 0.63, P = 1 \times 10^{-35} \); brown: \( r = 0.54, P = 1 \times 10^{-24} \); midnight blue: \( r = 0.34, P = 6 \times 10^{-10} \); Dataset S11). Among them, the red module contained GTR2, PEX14, and other genes related to the transport of some other metabolites. The brown module contained genes related to GSL synthesis, such as BATS and IPMI2 (AT2G43100). At 40 DAF, two modules were significantly positively correlated with SGC (black: \( r = 0.69, P = 2 \times 10^{-39} \); purple: \( r = 0.34, P = 5 \times 10^{-9} \); Dataset S12). GTR2 and PEX14 are also present in the black module. Additionally, we found that the gene significance (GS) of these genes was very high in two stages (BnaC02.GTR2: GS20 DAF = 0.64, GS40 DAF = 0.55; BnaA09.PEX14 (BnaA09g06240D): GS20 DAF = 0.56, GS40 DAF = 0.56; Figure 2c, Datasets S13 and S14).

To identify the key genes controlling the SGC of B. napus, we extracted 74 genes significant in TWAS localized in SGC-related QTL intervals at two stages. Of these 74 genes, 39 genes, including the SGC-related genes orthologous to GTR2, HAG1/ MYB28, APO3 (AT5G61930), TAF5 (AT5G25150), PYRD (AT5G23300), and ORE14 (AT5G62000), were present in 15 stable QTLs (Figure 2c; Dataset S15). The POCKET algorithm, which integrates multi-omics data and is based on machine learning, can prioritize candidate genes in QTL intervals (Tang et al., 2020). We used POCKET to analyse the functional ranking of co-localized genes in qGSL.C02.1. BnaC02.GTR2 with the highest GS at 20 and 40 DAF was highly ranked in qGSL.C02.1 (Figure S5, Dataset S16). To study the effect of BnaC02.GTR2 on SGC accumulation, we analysed the relationship between variant genotypes and phenotypes of BnaC02.GTR2 in the population. The coding region and the upstream 2 kb promoter region of BnaC02.GTR2 were classified into four major haplotypes in 505 accessions, and a significant difference in SGC between the haplotypes (Figure 2d) was observed. The haplotype B possessed a significantly higher average SGC (94.40 μmol/g) than the other three haplotypes with a P-value of 5.71 × 10^{-17}, whereas no significant difference in average SGC among the remaining three haplotypes was observed. Therefore, we speculated that sequence variations in the BnaC02.GTR2 coding and promoter regions might be responsible for variation in SGC in B. napus. In addition, correlation analysis of population phenotype and gene expression showed that BnaC02.GTR2 positively regulated the

![Figure 2](image-url)
SGC in *B. napus* \( R^2_{\text{DAF}} = 0.40, \ C^2_{\text{DAF}} = 2.85 \times 10^{-22}; \ R^2_{\text{DAF}} = 0.28, \ C^2_{\text{DAF}} = 1.44 \times 10^{-32}; \) \( R^2_{\text{DAF}} = 0.40, \ P_{\text{DAF}} = 2.85 \times 10^{-32}; \) \( R^2_{\text{DAF}} = 0.28, \ P_{\text{DAF}} = 1.44 \times 10^{-21}; \) \( R^2_{\text{DAF}} = 0.40, \ P_{\text{DAF}} = 2.85 \times 10^{-32}; \) \( R^2_{\text{DAF}} = 0.28, \ P_{\text{DAF}} = 1.44 \times 10^{-21}; \) These results indicate that *BnaC02.GTR2* could be the causal gene underlying *qGSL.C02.1*.

**Selection analysis of QTLs associated with low SGC in *B. napus* breeding**

To explore the QTLs that could be utilized in breeding programs to develop low-SGC varieties, we divided the lead SNP of QTLs into ancient and derived haplotypes, of which the latter generated from the ancient alleles (Tang et al., 2020). We found that nine QTLs (*qGSL.A02.1*, *qGSL.A02.2*, *qGSL.A08.1*, *qGSL.A09.1*, *qGSL.A09.2*, *qGSL.A09.4*, *qGSL.C02.1*, *qGSL.C09.1*, and *qGSL.C09.3*) had specific haplotypes. Comparison of the nucleotide diversity \( \pi \) values of ancient and derived haplotypes revealed that six QTLs were strongly selected and four of these were located on chromosomes A09 and C09 (Figures 3a,c,e, and S6). Based on the structural delineation of 505 accessions in previous studies, semi-winter population 1 (SW1) corresponded to low SGC and low-erucic acid accessions and semi-winter population 2 (SW2) corresponded to double-high accessions (Tang et al., 2020). In different subpopulations, the derived alleles of *qGSL.A09.2* and *qGSL.C09.1* had major effects on SGC (Figures 3d,f, and S7A to I). Based on the results of QTL candidate gene prediction, *BnaA09.PYRD* and *BnaC09.HAG1/MYB28* were predicted to be the candidate genes of *qGSL.A09.2* and *qGSL.C09.1*, respectively (Figures S8 and S9).

![Figure 3](https://example.com/figure3.png)

**Figure 3** Analysis of nucleotide diversity (a) of three SGC-associated QTLs. \( \pi \) of the accession groups and the GWAS signal of the mixed linear model (MLM) in (a) *qGSL.C02.1*, (c) *qGSL.C09.1*, and (e) *qGSL.A09.2*. Boxplot for SGC between two types of haplotypes in (b) *qGSL.C02.1*, (d) *qGSL.C09.1*, and (f) *qGSL.A09.2*. (g) Venn diagram showing overlapped low-SGC materials with low-SGC haplotypes in *qGSL.C02.1*, *qGSL.C09.1*, and *qGSL.A09.2*. (h) Analysis of different QTL haplotypes in representative materials. The first row is SGC and the second row is material. ‘+’ indicates that the allele is present, ‘−’ indicates that the allele is absent. In (a), (c), and (e), the accession groups are divided by the genotype of the lead SNP in QTLs. The dots in the upper panel are sliding regions of \( \pi \) (20 kb window sliding in 10 kb steps) by the two accession groups, and the curves are lower smoothed dots. Colours in the GWAS signal indicate the LD \( R^2 \) between the lead SNP and each variant.
BnaC02.GTR2 and its paralog BnaA02.GTR2 (BnaA02g33530D) were present in qGSL.C02.1 and qGSL.A02.2, respectively. Interestingly, the ancient and derived haplotypes of the two QTLs had low and similar \( p \) values (Figures 3a and S6D). This suggests that GTR2 may have indispensable functions during seed development, making it highly conserved in B. napus. Both qGSL.C02.1 and qGSL.A02.2 had large effects on SGC in subpopulations (SW1 and SW2), and their low-SGC haplotypes were mainly distributed in SW1 with low-SGC phenotypes (Figures 3g and S7A,B). Interestingly, the ancient and derived haplotypes of the two QTLs based on haplotype analysis (qGSL.A02.2; \( p = 0.40 \), \( \rho_{TSW} = 0.42 \); qGSL.C02.1; \( p = 0.46 \), \( \rho_{TSW} = 0.08 \); Figure S10). Taken together, these results suggest that despite the conservation status of both BnaA02.GTR2 and BnaC02.GTR2, breeding selection strategies effectively utilized the natural variations that singularly affect SGC.

Notably, the phenotypic distribution of haplotypes for the four QTLs (qGSL.A02.2, qGSL.C02.1, qGSL.A09.2, and qGSL.C09.1) were similar with the mean difference of SGC between the two haplotypes ranging from 52.60 to 67.47 μmol/g in the population (Figures 3b,d,f and S6D). To explore how the phenotypic distribution characteristics of these QTLs were shaped, we analysed the haplotypes of qGSL.A09.2, qGSL.C01.9, and qGSL.C02.1 that represent the effects of GTR2 in both qGSL.C02.1 and qGSL.A02.2 in the 154 low-SGC accessions. These results showed that in 151 accessions, the low-SGC haplotypes were identically associated with all three QTLs (Figure 3g). We then selected the top five accessions with extremely low SGC and extremely high SGC from the 505 accessions, respectively. Notably, the low-SGC haplotypes in the four large-effect QTLs (qGSL.A02.2, qGSL.C02.1, qGSL.A09.2, and qGSL.C09.1) were consistently present in the five low-SGC accessions (Figure 3h). These results suggested that the low-SGC accessions were mainly generated by co-selection of multiple large-effect QTLs, of which contained key GSL-associated genes GTR2, PYRD, and HAG1/MYB28 (Figure 3a,c,e).

Disruption of Bna.GTR2s conferred low SGC but affected several seed traits in B. napus

The correlation coefficient analysis showed that six GTR2 homologs in B. napus could be divided into two groups. The first group, including BnaA06.GTR2 (BnaA06g22160D), BnaC03.GTR2 (BnaC03g51560D), BnaA02.GTR2 and BnaC02.GTR2 had a higher protein sequence similarity with GTR2 than the second group containing BnaA09.GTR2 (BnaA09g06190D) and BnaC09.GTR2 (BnaC09g05810D) (Figures 4A and S11A). Therefore, to explore the potential functions of Bna.GTR2 homologs, we used the CRISPR/Cas9 genome editing system to edit the four homologous genes belonging to the first group (Figure S11A). We designed two sgRNAs (sgRNA1 and sgRNA2) that specifically target the major facilitator superfamily conserved domain (Wang et al., 2020) in exons 4 and 5 of all four Bna.GTR2 homologs, respectively (Figures S11B,C). In total, 29 \( T_2 \)-positive transgenic plants were obtained and 10 homozygous \( T_2 \) lines with mutations in the four homologous genes were identified by sequencing the target sites. The homozygous mutations caused amino acid substitutions,

![Figure 4](image-url)  
*Figure 4* Functional characterization of BnaC02.GTR2 as a positive regulator of SGC in B. napus. (A) Protein sequence similarity analysis of GTR2 in B. napus and A. thaliana. (B) Boxplot of SGC (μmol/g) in WT and Bna.gtr2 mutants. (C) Boxplot of TSW (g) in WT and Bna.gtr2 mutants. (D) Boxplot of seed oil content (% of dry weight) in WT and mutants. In (B)–(D), letters a, b, c, d, e, and f indicate significant differences by two-way ANOVA followed by post hoc Tukey’s HSD test for all pairwise comparisons (\( P < 0.05 \)). a02, c02, a06, and c03 represent mutants of BnaA02.gtr2, BnaC02.gtr2, BnaA06.gtr2, and BnaC03.gtr2, respectively; a2c2, a2c2c3, and a2c2a6c3 represent double, triple, and quadruple mutants of Bna.gtr2. Each colour represents a material, and each dot represents a replicate. The middle bars represent the median, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. (E) Phenotypes of 50 seeds of WT and mutants, Bar = 1 cm. Seed diameters (mm) are the mean ± SD of five biological replicates.
frameshifts or premature stop codons that resulted in truncated MFS conserved domains (Dataset S17). Further sequence analysis showed that the potential off-target sites were not mutated (Datasets S17 and S19).

Subsequently, we measured the SGC phenotype of different Bna.grt2 homozygous mutants. Each of the four single-locus Bna.grt2 mutants showed significant decrease in SGC compared with the wild-type (WT) recipient (cv.13CK-3), with BnaA02.grt2 (a02) and BnaC02.grt2 (c02) showed the most significant decrease, whereas BnaC03.grt2 (c03) and BnaA06.grt2 (a06) showed a moderate decline. Moreover, the BnaA02.grt2 and BnaC02.grt2 double mutant (a2c2), as well as BnaA02.grt2, BnaC02.grt2 and BnaC03.grt2 triple mutants (a2c2c3) also exhibited significant decreases in SGC compared with a02 or c02. Notably, the average SGC of the BnaA02.grt2, BnaC02.grt2, BnaA06.grt2, and BnaC03.grt2 quadruple mutant (a2c2a6c3) was 18.21 μmol/g, representing an 86.85% decrease compared with WT (Figure 4B; Dataset S20). These results revealed the crucial function of Bna.GTR2s in seed GSL accumulation. In addition, our findings showed that different Bna.GTR2 homologs studied here affected SGC to different degrees and a partially functional redundancy commonly existed among the four homologs.

Surprisingly, the TSW of the a2c2a6c3 quadruple mutant was decreased by 64.35% compared with that of WT seeds, similar to the reduction degree of SGC (Figure 4C; Dataset S21). Additionally, seed diameter showed significant decrease in Bna.grt2 mutants compared with WT seeds (Figures 4E and S12, Dataset S22). Previous studies have shown that the main storage compounds accumulated in seeds of B. napus consist of storage triacylglycerols and proteins (Schwender et al., 2006). To investigate whether Bna.GTR2s affect other seed traits of B. napus, we examined SOC, protein content, and fatty acid composition in the mutants and WT seeds. Interestingly, the Bna.grt2 mutant seeds showed a significant increase in SOC, a significant change in fatty acid composition, and varying degrees of decrease in protein content (Figures 4D and S13, Dataset S20). These results suggest that Bna.GTR2s can affect not only the SGC but also seed size and may participate in multiple aspects of dry matter accumulation in B. napus seeds.

GSLS are synthesized mainly in vegetative tissues, such as leaves and roots, and are finally transported to seeds in Arabidopsis. Therefore, we also detected the GSL content in the roots and leaves of BnaC02.grt2 mutants. The results showed that the GSL content of the roots and the leaves of the mutants were 9.42 and 6.29 μmol/g, respectively. Compared with WT, the GSLs content of the mutants were significantly decreased in the leaves but significantly increased in the roots, which means that the transport of GSL synthesized in the roots to other vegetative tissues may be inhibited, resulting in the accumulation of GSL in the roots. (Figure S14, Dataset S23).

**BnaC02.GTR2 is involved in multiple processes during the seed development of B. napus**

It is noteworthy that the altered SGC of BnaC02.grt2 mutant was accompanied by a variety of traits associated with seed development (Figures 4B to E), which means that SGC-related genes or BnaC02.GTR2 may be associated with seed development. To verify this hypothesis, we first re-observed the genes detected by TWAS at 20 and 40 DAF that were significantly associated with SGC and found that 73 and 56 of them were associated with amino acid, sugar, and fatty acid accumulation in Arabidopsis seeds, respectively (Figures S15a,b, Datasets S24 and S25).

These results suggested that SGC-related genes were inextricably linked to the process of seed development. Then, we conducted a correlation analysis between the expression of BnaC02.GTR2 and genes significant in TWAS related to seed development at 20 DAF [the critical period of seed dry matter accumulation (Yu et al., 2010)]. The results showed that BnaC02/GTR2 was significantly positively correlated with genes, such as ARF2 (AT5G62000), EMB204 (AT5G24400), etc., and negatively correlated with genes, such as SWEET5 (AT5G62850), TD4 (AT5G39300), and UGT71I1 (AT4G22772). These genes are involved in multiple seed developmental processes (Figure 5a, Dataset S26; Bock et al., 2003; Engel et al., 2005; Meinke, 2020; Okushima et al., 2005; Schruff et al., 2006; Tzafrir et al., 2004; Wang et al., 2008; Yu et al., 2005). Next, we constructed a co-expression analysis of BnaC02.GTR2 and seed development-related genes identified by TWAS at both 20 and 40 DAF and found that BnaC02.GTR2 is associated with multiple processes, such as plant development, protein accumulation, amino acid synthesis, GSL synthesis, sugar assimilation, and endosperm development (Figure 5b; Dataset S26). In addition, we used the gene expression of BnaC02.GTR2 at 20 DAF as a covariate for TWAS to re-performed TWAS analysis. Four hundred and six genes disappeared significantly compared with the previous result, and these genes were significantly enriched in the synthesis of GSL and different pathways of seed development (Dataset S27, Figure S16). These results suggested that BnaC02.GTR2 is an important factor affecting genes related to seed development. And we also analysed the transcriptome of WT and BnaC02.grt2 mutant in developing seeds at 20 DAF. We identified 45 076 expressed genes (mean TPM > 1 for each type), including 1234 were differentially expressed (log2foldchange > 1 and Padj < 0.05). A significant decrease in the expression of genes associated with GSL synthesis and transport was observed. Interestingly, not only is the expression of genes regulating aliphatic GSL [such as HAG1/MYB28, CYP83A1, BAT5, AOP2 (AT4G03060) significantly reduced, but also for the synthesis of indolic GSL [e.g. CYP91F1 (AT4G37430), GSTF9 (AT2G30860), and IGMT1 (AT1G21100), suggesting possible crosstalk between GSL transport and synthesis pathways. And we also observed a significant reduction in the expression of indolic GSL transporter GTR3. These implied that aliphatic and indolic GSL transporters may cooperate for GSL transport (Figure 5c; Datasets S28 and S29). Moreover, the expression of genes related to seed development also changed significantly [such as MYC4 (AT4G17880), CYP78A9 (AT3G61880), and CCAOMT7 (AT4G26220)]. Among them, MYC4 plays a redundant role with MYC3 and MYC2 in the accumulation of Arabidopsis seed storage protein (Gao et al., 2016), which might be related to the protein content alteration in the seeds of BnaC02.grt2 mutant (Figure S13A). Cytochrome P450 CYP78A9 is involved in Arabidopsis reproductive development and also regulates silique length and seed weight in B. napus (Sotelo-Silveira et al., 2013). In B. napus, a CACTA-like transposable element in the upstream region of BnaA9.CYP78A9 acts as an enhancer to increase silique length and seed weight (Shi et al., 2019). CCAOMT7 encodes a caffeoyl-coenzyme A O-methyltransferase (CCAOAMT)-like protein, which contributes to lignin biosynthesis in the cell wall and may have a role on the development of seed coat (Yang et al., 2019; Yu et al., 2009). Furthermore, GO enrichment analysis indicated that the differentially expressed gene set between BnaC02.grt2 mutant and WT at 20 DAF was enriched in multiple biological processes, including
sulphate assimilation, GSL biosynthesis, carbohydrate metabolism, lipid binding, transporter activity, amino acid binding, and oligopeptide transport (Figure 5d; Dataset S30).

Accordingly, these findings strongly suggest that SGC-related genes have possible associations with seed development. BnaC02.GTR2, with pleiotropic functions, might not only play a role in seed GSL accumulation, but also be involved in multiple approaches of seed development in B. napus.

Discussion

GSLs are secondary metabolites distributed specifically in cruciferous plants (Halkier and Gershenzon, 2006; Sonderby et al., 2010). Although GSL has a very weak effect on the quality of crushed rapeseed oil, they can largely determine the commercial value of crushed meals of Brassica crops (Bisht and Augustine, 2019; Ishida et al., 2014; Kuznetsova and Polyakova, 2019). However, it is challenging to further reduce the SGC to <15 µmol/g dry meal in B. napus, despite persistent efforts made in the past half-century, which greatly impedes the utilization of rapeseed meal as an alternative to soybean meal. In this study, we successfully dissected the genetic basis of SGC in B. napus through an integrative strategy by combining the GWAS and TWAS approaches. Compared with previous studies, our GWAS analysis showed high accuracy and reliability in identifying the loci controlling SGC—the SGC-QTLs were consistent with the QTLs identified on chromosomes A02, A09, C02, C07, and C09 reported before (Cun-Min et al., 2015; Feng et al., 2012; Li et al., 2014; Liu et al., 2020). In addition, several novel QTLs, such as qGSL.C03.1 and qGSL.A09.4 (Figure 1a; Dataset S2), were also identified.

Figure 5 BnaC02.GTR2 possesses pleiotropic effects during seed development. (a) Correlation analysis of BnaC02.GTR2 and genes related to seed development. Red circles represent positive correlations with the expression of BnaC02.GTR2. Blue circles represent negative correlations with the expression of BnaC02.GTR2 and grey circles represent no correlation. Light green lines indicate the correlation between other genes. (b) Co-expression analysis of BnaC02.GTR2 at 20 and 40 DAF. Circles represent genes. Squares represent the biological process; shape size represents the absolute value of the correlation between BnaC02.GTR2 and the genes in biological processes. (c) Volcano plot of BnaC02.gtr2 mutant. The Y-axis is the adjusted P-value, and the X-axis is log2 fold-change (FC) between mutant and WT. Grey lines are at log2FC = 1 or adjusted P-value = 0.05. The significant genes in TWAS for SGC at 20 DAF are labelled with red points with gene names. (d) Enrichment analysis with differentially expressed gene set between BnaC02.gtr2 mutant and WT at 20 DAF.
Seed development at 20 and 40 DAF is important for seed expansion, embryo development and accumulation of protein and GSL in *B. napus* (Dong et al., 2004; Larden and Trboi-Blondel, 1994; Yu et al., 2010). Combining the results from TWAS, WGCNA, and POCKET, we predicted the potential candidate genes underlying the SGC-QTLs and revealed some key genes involved in seed GSL accumulation. We identified 324 overlapping genes that were significantly associated with SGC in the two seed development stages. These genes significant in TWAS are likely to play a role in the regulation of SGC directly or indirectly. Seventy-four GWAS and TWAS co-localized genes, such as *HAG1/MYB28* (Sonderby et al., 2007), *BCAT4* (AT3G19710; Schuster et al., 2006), and *MAM1* (AT5G23010; Textor et al., 2004) involved in diverse dimensions of GSL synthesis and accumulation and the genes responsible for GSL transport such as GTR2 (Nour-Eldin et al., 2012) were identified. Overall, these results based on integrative mega data analysis, for the first time, constructed a comprehensive network including QTLs and co-expressed genes involved in the regulation of SGC. These findings can assist in unravelling the genetic architecture of GSL synthesis and transport and exploring the evolution of gene function in related *Brassica* plants.

Since low SGC is a major breeding objective of *B. napus* for the past decades, we performed a selection analysis of the QTLs associated with SGC. As expected, most low-SGC haplotypes in these QTLs were selected in modern ‘double-low’ varieties. Furthermore, the haplotype correlation analysis of loci in high and low SGC accessions demonstrated a clear co-selection of multiple QTLs, and most low-SGC accessions were produced mainly by selecting three QTLs—qGSL.C02.1, qGSL.A09.2, and qGSL.C09.1 (Figure 3g,h and S6). As evident from the breeding history of low-SGC varieties, the reduced concentrations of GSL are often a consequence of crossing from the breeding history of low-SGC varieties, the reduced concentrations of GSL in natural accessions should be carefully evaluated in the future. Furthermore, we verified that all four *Bna GTR* homologs contributed to the regulation of SGC, and their effects varied among different copies. Consistent with the evidence in Arabidopsis (Nour-Eldin et al., 2012), *B. napus* and *B. juncea* (Nour-Eldin et al., 2017), our results clearly indicated the possibility of using GTRs to exclusively regulate SGC in *B. napus* (Figure 4b).

However, investigation of important seed traits of mutants indicated the pleiotropic roles of *Bna GTR*2s during seed development, affecting other important seed traits, such as TSW and SOC. This speculation was subsequently confirmed by transcriptome analysis of mutants (Figure 5c,d), TWAS, and co-expression analysis (Figure 5a,b), in which *Bna C02 GTR2* with genes involved in multiple biological processes, such as amino acid accumulation, GSL synthesis and transport, sugar assimilation, and oil accumulation.

From an agricultural perspective, the utilization of source-sink transport engineering for GTR transporters is expected to serve as a new strategy to address the challenges of oilseed breeding (Nour-Eldin et al., 2012). Although the selection analysis revealed that the genomic regions flanking *Bna C02 GTR2* displayed low nucleotide diversity, the haplotypes leading to low-SGC were selected during breeding because of their significant effect on the reduction of SGC (Figure 3a,b). Moreover, different unlike CRISPR mutants of *Bna C02 gtr2*, the natural low-SGC haplotype had no significant effect on seed traits, such as TSW and SOC (Figures 6 and S10). Based on these findings, we conjectured that the low-SGC haplotype of *BnaC02 GTR2* was utilized in the breeding selection process that singularly affected SGC in *B. napus*. In other words, the function of *Bna C02 GTR2* might have been modified partially by the selection process; therefore, even though the function of *Bna C02 GTR2* and the low-SGC *Bna C02 GTR2* allele had a reduced GSL transportation activity, it was still able to maintain the normal development and metabolism of the seeds. These findings suggest that the functional difference between the natural variations and CRISPR mutants of *Bna C02 gtr2* might depend on the structural alteration of the gene.

The GTR transporters belong to the ubiquitous NRT/PTR family (NPF) and may have evolved from the original cyanogenic glucoside transporter transporting either cyanogenic glucosides or GSL (Jorgensen et al., 2017). Structural studies of bacterial PTSs and NPF6.3 (Aduri et al., 2015; Doki et al., 2013; Parker and Newstead, 2014; Solcan et al., 2012; Sun et al., 2014) suggested the possible roles of amino acid residues P7 and P9 in the interaction with glucose (Jorgensen et al., 2017). However, whether these residues can interact with GSL and other vital metabolites with similar structures indiscriminately, thereby affecting SGC and seed development, remains unclear. Moreover, knowledge of the gene structure, substrate binding specificity and regulatory mechanisms of GTR transporters remain to be explored in *B. napus*. Taken together, our results provide a promising strategy for further application of GTR transporters to specifically reduce the SGC in *B. napus*, such as the precise modification by gene editing.
Methods

Plant materials and collection of phenotype data

A total of 505 *B. napus* accessions (Tang et al., 2020) collected from different research institutions were grown in Wuhan for 3 years (2017–2019). Leaves at seeding stage were sampled and quickly frozen in liquid nitrogen. Mature open-pollinated seeds of the natural population comprising 505 *B. napus* inbred lines were harvested and desiccated to analyse SGC and other traits (SOC, seed protein content, and seed fatty acids) by a Foss NIRSystems 5000 near-infrared reflectance spectroscope (Gan et al., 2003). Six replications were used for the analysis of every accession. The measurements were carried out at the National Research Center of Rapeseed Engineering and Technology, Huazhong Agricultural University, Wuhan, China.

Variation calling and annotation

The genome sequences of *B. napus* were downloaded from GenBank (http://www.genoscope.cns.fr/brassicanapus/, Chalhoub et al., 2014). Readings from resequencing (Tang et al., 2020) were compared with the genome using the parameter ‘mem -M -k 32 -t 4’ via the BWA software (Li, 2013). Then, PCR duplicates of sequencing reads were removed using SAMtools and mapped reads were retained in the BAM file format (Li et al., 2012). For genotyping, we applied the previous projections (Tang et al., 2020). GATK v3.6 (McKenna et al., 2010) was used to identify the existing sequence variations in 505 *B. napus* accessions and to merge the files. The SNPs and InDels with lower mapping quality (MQ < 20) or reads with lower sequencing depth (DP < 50) were also filtered. After obtaining the genotyping data, the LD-KNN algorithm was used to estimate the missing genotype (Chen et al., 2014), and the accuracy of the data was ensured at >99.7%.

GWAS for SGC

A total of 10 620 048 high-quality SNPs (MAF > 0.05) in all populations were used for GWAS and the significance thresholds for association were calculated using the Genetic Type I Error Calculator (GEC) software (Li et al., 2012) as 7.96 × 10^-7. We used the GEMMA software package (Zhou and Stephens, 2012) to correlate the 3-year WH SGC trait data using a mixed linear model. The POCKET from the GitHub (https://github.com/zhaouu/POCKET; Tang et al., 2020) was downloaded and combined multiple omics data, large-effect variations, haplotype analysis, and known trait-associated functional genes to prioritize candidate genes in SGC-related QTLs of *B. napus*.

Weighted gene co-expression network analysis

To study the significantly associated gene co-expression profiles of GWAS and TWAS at 20 and 40 DAF, the expression of genes at each period was analysed by the WGCNA method (Langfelder and Horvath, 2008). The expression matrices were constructed, genes with similar expression patterns were clustered into the

Figure 6 A putative low-SGC breeding selection model. The dashed line indicates complete loss of function of GTR2 is not utilized, and the solid line indicates the process of breeding selection. An oval divided in half indicates a loss of function of BnaC02.GTR2. The dashed box indicates the processes potentially affected by BnaC02.GTR2. Boxplot indicates the SGC in low-SGC SW1 subpopulation corresponding to two major haplotypes of four loci selected during breeding.
same module, and correlation analysis was performed with SGC. Pearson correlation coefficients were calculated for the correlation analysis. The NetworkType and TOMType parameters are set to ‘signed’. The modules were divided using the following parameters: deepSplit – 4, miniModuleSize – 20, and mergeCutHeight – 0.1. The 20 and 40 DAF soft power were set at 14 and 8, respectively. Finally, correlation analysis was performed using Cytoscape v3.5.0 software for thioide, and the module and gene correlation networks were visualized and displayed.

**TWAS for SGC**

RNA-seq data were analysed for quality control using FastQC (Ewels et al., 2016). STAR software (Dobin et al., 2013) was used to compare filtered reads with the *B. napus* reference genome, quantitative analysis of gene expression using Salmon (Patro et al., 2017), and standardization using tximport (Charlotte et al., 2016). Correlation analysis was performed using the EMMAX software (Kang et al., 2010), as well as as inline and LMMs. SNPs were randomly selected from the genome to calculate the kinship matrix. The FDR-corrected P-value of ≤ 0.05 was used as the significance threshold for TWAS analysis.

**GO enrichment analysis**

All *B. napus* genes were compared with all *A. thaliana* proteins using BLASTP (Jacob et al., 2008). The most important proteins in *A. thaliana* (probability threshold 10\(^{-5}\) – 1) were selected for enrichment with the *A. thaliana* genome. Enrichment analysis was performed using Fisher’s exact test, and the OR and P values were calculated.

**Construction of the CRISPR/Cas9 vector and plant transformation**

The CRISPR/Cas9 genome editing system, used in this study to edit the *BnaC02* genes was developed following previous methods (Xing et al., 2014). The primers used for the construction of the sgRNA vector are listed in the Dataset S18. The hypocotyl of the sgRNA vector are listed in the Dataset S18. The hypocotyl were generated. Quality control of RNA-seq reads was performed using the NEBNext UltraTM RNA Library Prep Kit for Illumina® (NEB, USA). The library preparations were sequenced on an Illumina HiSeq 4000 platform and 150 bp paired-end reads were generated. Quality control of RNA-seq reads was performed using FastQC software and summarized using multi-qc (Ewels et al., 2016). We used Salmon (Patro et al., 2017) to quantify the reads of annotated genes in the *B. napus* genome (http://www. genomescope.cns.fr/brassicanapus/), which were then imported and normalized with tximport (Sonesson et al., 2015) and DESeq2 was used for differential expression analysis (Love et al., 2014).

**Acknowledgements**

This research was supported by the National Key Research and Development Program of China (2016YFD0101000), the National Natural Science Foundation of China (31571704), the Natural Science Foundation of Hubei (2019CFA090), the Wuhan Applied Foundational Frontier Project (2019020701011446), and the 111 Project (B20051).

**Conflict of interest**

The authors declare no competing interests.

**Author contributions**

H.D. and L.G. designed and supervised the study. Z.T., Y.Z., and H.Z. performed the bioinformatics analysis. Z.T., Z.X., L.D., and Y.Z. prepared the manuscript. D.H., L.G., and X.Y. revised the manuscript. All the authors have read and approved the manuscript.

---

**Genetic basis of seed glucosinolate content in rapeseed**

same module, and correlation analysis was performed with SGC. Pearson correlation coefficients were calculated for the correlation analysis. The networkType and TOMType parameters are set to ‘signed’. The modules were divided using the following parameters: deepSplit – 4, miniModuleSize – 20, and mergeCutHeight – 0.1. The 20 and 40 DAF soft power were set at 14 and 8, respectively. Finally, correlation analysis was performed using Cytoscape v3.5.0 software for thioide, and the module and gene correlation networks were visualized and displayed.

**TWAS for SGC**

RNA-seq data were analysed for quality control using FastQC (Ewels et al., 2016). STAR software (Dobin et al., 2013) was used to compare filtered reads with the *B. napus* reference genome, quantitative analysis of gene expression using Salmon (Patro et al., 2017), and standardization using tximport (Charlotte et al., 2016). Correlation analysis was performed using the EMMAX software (Kang et al., 2010), as well as as inline and LMMs. SNPs were randomly selected from the genome to calculate the kinship matrix. The FDR-corrected P-value of ≤ 0.05 was used as the significance threshold for TWAS analysis.

**GO enrichment analysis**

All *B. napus* genes were compared with all *A. thaliana* proteins using BLASTP (Jacob et al., 2008). The most important proteins in *A. thaliana* (probability threshold 10\(^{-5}\) – 1) were selected for enrichment with the *A. thaliana* genome. Enrichment analysis was performed using Fisher’s exact test, and the OR and P values were calculated.

**Construction of the CRISPR/Cas9 vector and plant transformation**

The CRISPR/Cas9 genome editing system, used in this study to edit the *BnaC02* genes was developed following previous methods (Xing et al., 2014). The primers used for the construction of the sgRNA vector are listed in the Dataset S18. The hypocotyl of the sgRNA vector are listed in the Dataset S18. The hypocotyl were generated. Quality control of RNA-seq reads was performed using the NEBNext UltraTM RNA Library Prep Kit for Illumina® (NEB, USA). The library preparations were sequenced on an Illumina HiSeq 4000 platform and 150 bp paired-end reads were generated. Quality control of RNA-seq reads was performed using FastQC software and summarized using multi-qc (Ewels et al., 2016). We used Salmon (Patro et al., 2017) to quantify the reads of annotated genes in the *B. napus* genome (http://www. genomescope.cns.fr/brassicanapus/), which were then imported and normalized with tximport (Sonesson et al., 2015) and DESeq2 was used for differential expression analysis (Love et al., 2014).

**Acknowledgements**

This research was supported by the National Key Research and Development Program of China (2016YFD0101000), the National Natural Science Foundation of China (31571704), the Natural Science Foundation of Hubei (2019CFA090), the Wuhan Applied Foundational Frontier Project (2019020701011446), and the 111 Project (B20051).

**Conflict of interest**

The authors declare no competing interests.

**Author contributions**

H.D. and L.G. designed and supervised the study. Z.T., Y.Z., and H.Z. performed the bioinformatics analysis. Z.T., Z.X., L.D., and Y.Z. prepared the manuscript. D.H., L.G., and X.Y. revised the manuscript. All the authors have read and approved the manuscript.
References
Aduri, N.G., Prabhala, B.K., Ernst, H.A., Jorgensen, F.S., Olsen, L. and Mirza, O. (2015) Salt bridge swapping in the EXD2RXFYY motif of proton-coupled oligopeptide transporters. J. Biol. Chem. 290, 29931–29940.
Agerbirk, N. and Olsen, C.E. (2012) Glucosinolate structures in evolution. Phytochemistry, 77, 16–45.
Agnihotri, A., Prem, D. and Gupta, K. (2007) The chronicles of oil and meal quality improvement in oilseed rape. Adv. Bot. Res. 45, 49–97.
Andersen, T.G., Nour-Eldin, H.H., Fuller, V.L., Olsen, C.E., Burow, M. and Halkier, B.A. (2013) Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative Arabidopsis. Plant Cell, 25, 3133–3145.
Bak, S. and Feyerisen, R. (2001) The involvement of two P450 enzymes, CYP83B1 and CYP83A1, in aconitosemethylase and glucosinolase biosynthesis. Plant Physiol. 127, 108–118.
Barco, B. and Clay, N.K. (2019) Evolution of glucosinolate diversity via whole-genome duplications, gene rearrangements, and substrate promiscuity. Annu. Rev. Plant Biol. 70, 585–604.
Beekwilder, J., van Leeuwen, W., van Dam, N.M., Bertossi, M., Grandi, V., Miczi, L., Soloviev, M. et al. (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. PLoS ONE, 3, e2608.
Bisht, N.C. and Augustine, R. (2019) Development of Brassica Olleid Seeds with Low Antinutritional Glucosinolates and Rich in Anticancer Glucosinolates: Nutritional Quality Improvement in Plants (pp. 271–287). Switzerland: Springer, Cham.
Blazevic, I., Montaut, S., Burcul, F., Olsen, C.E., Burow, M., Rollin, P. and Agerbirk, N. (2020) Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. Phytochemistry, 169, 112–100.
Bock, K.W., Honys, D., Ward, J.M., Padmanabhan, S., Nawrocki, E.P., Hirschi, K.D., Twell, D. et al. (2006) Integrating membrane transport with male gametophyte development and function through transcriptomics. Plant Physiol. 140, 1151–1168.
Brondani, C., Rangel, F.H.N., Brondani, R.P.V. and Ferreira, M.E. (2002) QTL mapping and introgression of yield-related traits from Orzya glumaepatula to cultivated rice (Oryza sativa) using microsatellite markers. Theor. Appl. Genet., 104, 1192–1203.
Brown, P.D., Tokuhisa, J.G., Reichelt, M. and Gershenson, J. (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry, 62, 471–481.
Burow, M., Halkier, B.A. and Klebeinstein, D.J. (2010) Regulatory networks of glucosinolates shape Arabidopsis thaliana fitness. Curr. Opin. Plant Biol. 13, 348–353.
Burton, W., Salisbury, P. and Potts, D. (2003a) Inheritance of allyl glucosinolate in the development of canola quality Brassica juncea for Australia. In Proc. 11th Int. Rapeseed Congr., p. 280.
Burton, W., Salisbury, P. and Potts, D. (2003b) The potential of canola quality Brassica juncea as an oilseed crop for Australia. In Proceedings 13th Biennial Australian Research Assembly on Brassicas, Tamworth, NSW, pp. 62–64.
Buss, C.E., Tizioto, P.C., Oliveira, P.S.N., Mudalud, M.A., Cesar, A.S.M., Ventura, R.V., Afonso, I. et al. (2016) O891 Genome-wide efficient mixed-model study for meat quality in Nellore cattle. J. Anim. Sci. 94(suppl_5), 428–429.
Celenza, J.L., Quiel, J.A., Smolen, G.A., Merrikh, H., Silvestro, A.R., Normanly, J. and Bender, J. (2005) The Arabidopsis ATR1 Myb transcription factor controls gametophyte development and function through transcriptomics. F1000 Res. (suppl_5), 481.
Chen, S.X., Petersen, B.L., Olsen, C.E., Schulz, A. and Halkier, B.A. (2001) Long-distance phloem transport of glucosinolates in Arabidopsis. Plant Physiol. 127, 194–201.
Chen, W., Gao, Y.Q., Xie, W.B., Gong, L., Lu, K., Wang, W.S., Li, Y. et al. (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat. Genet. 46, 714–721.
Clarke, B.B. (2010) Glucosinolates, structures and analysis in food. Anal. Methods, 2, 310–325.
Cun-Min, Q., Shi-Meng, L., Xiu-Jian, D., Jin-Hua, F., Le-Dong, J., Hui-Yan, Z., Kun, L. et al. (2015) Identification of candidate genes for seed glucosinolate content using association mapping in Brassica napus L. Genes-Base, 6, 1215–1229.
DeLeo, M.V., Strahan, G.D., den Bakker, M. and Hoekenga, O.A. (2011) Weighted correlation network analysis (WGCNA) applied to the tomato fruit control of microRNA. PLoS ONE, 6, e26687.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P. et al. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29, 35–19.
Doki, S., Kato, H.E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., Iwase, N. et al. (2013) Structural basis for dynamic mechanism of proton-coupled transport of seed-abundant cDNAs. Plants, 218, 483–491.
Du, L. and Halkier, B.A. (1998) Biosynthesis of glucosinolates in the developing siliques walls and seeds of Sinapis alba. Phytochemistry, 48, 1145–1150.
Engel, M.L., Davis, R.H. and McCormick, S. (2005) Green sperm. Identification of male gamete promoters in Arabidopsis. Plant Physiol. 138, 2124–2133.
Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) MultiQIC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32, 3047–3048.
Fahey, J.W., Zalckalm, A.T. and Talalay, P. (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry, 56, 5–51.
Feng, J., Long, Y., Shi, L., Shi, J.Q., Barker, G. and Meng, J.L. (2012) Characterization of metabolite quantitative trait loci and metabolic networks that control glucosinolate concentration in the seeds and leaves of Brassica napus. New Phytol. 193, 96–108.
Freigmann, H. and Gogolashvili, T. (2014) MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in Arabidopsis thaliana. Mol. Plant, 7, 814–824.
Fu, Y., Lu, K., Qian, L.W., Mei, J.Q., Wei, D.Y., Peng, X.H., Xu, X.F. et al. (2015) Development of genic cleavage markers in association with seed glucosinolate content in canola. Theor. Appl. Genet. 128, 1029–1037.
Gan, L., Sun, X., Jin, L., Wang, G., Xu, J., Wei, Z. and Fu, T. (2003) Establishment of math models of NIRS analysis for oil and protein contents in seed brassica napus. Zhongguo Nong Ye Ke Xue, 36, 1609–1613.
Gao, C., Qi, S., Liu, K., Li, D., Jin, C., Li, Z., Huang, G. et al. (2016) MYC2, MYC3, and MYC4 function redundantly in seed storage protein accumulation in Arabidopsis. Plant Physiol. Biochem. 108, 63–70.
Gatrell, S., Lum, K., Kim, J. and Lee, X.G. (2014) Nonunimrnant nutrition symposium: potential of defatted microalgae from the biofuel industry as an ingredient to replace corn and soybean meal in swine and poultry diets. J. Anim. Sci. 92, 1306–1314.
Giamousta, A. and Mithen, R. (1995) The effect of modifying the glucosinolate content of leaves of oilseed rape (Brassica-Napus Sipolifea) on its interaction with specialist and generalist pests. Ann. Appl. Biol. 126, 347–363.
Gigolashvili, T., Berger, B., Mock, H.P., Muller, C., Weisshaar, B. and Flugge, U.I. (2007) The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. Plant J. 50, 886–901.
Gigolashvili, T., Berger, B. and Flugge, U. (2009) Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3-MYB transcription factors in Arabidopsis thaliana. Phytochem. Rev. 8, 3–13.
Gijzen, M., Mcgregor, I. and Seguin-Swartz, G. (1998) Glucosinolate uptake by Gasterophilus intestinalis. New Phytol., 140, 573–581.
}
Genetic basis of seed glucosinolate content in rapeseed

Hall, C., McCallum, D., Prescott, A. and Mithen, R. (2001) Biochemical genetics of glucosinolate modification in Arabidopsis and Brassica. Theor. Appl. Genet. 102, 369–374.

Hansen, C.H., Wittstock, U., Olsen, C.E., Hick, A.J., Pickett, J.A. and Hawk, B.A. (2001) Cytochrome p450 CYP79F1 from arabidopsis catalyzes the conversion of dihomomethionine and dihomothionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. J. Biol. Chem., 276, 11078–11085.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C. et al. (2012) Associative transcriptomics of traits in the polyplid crop species Brassica napus. Nat. Biotechnol. 30, 798–802.

Hein, O., Schwend, M. and König, W. (2006) Scale-free networks. WIRTSCHAFTSFORSCHUNG, 48, 267–275.

Hemm, M.R., Ruegger, M.O. and Chapple, C. (2003) The Arabidopsis ref2 mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. Plant Cell, 15, 179–194.

Howell, P.M., Sharpe, A.G. and Lydiard, D.J. (2003) Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (Brassica napus). Genome, 46, 454–460.

Ishida, M., Hara, M., Fukino, N., Kakizaki, T. and Morimitsu, Y. (2014) Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. Breed Sci. 64, 48–59.

Jacob, A., Lancaster, J., Buhler, J., Harris, B. and Chamberlain, R.D. (2008) Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways. PLoS. Genet. 9, e1003840.

Larden, A. and Triboi-Blondel, A.M. (1994) Freezing injury to ovules, pollen and seeds in winter rape. Br J. Exp. Physiol. 77, 77–80.

Liu, M., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

McKenna, A., Hanna, M., Banks, E., Svachenko, A., Cibulskis, K., Kernytsky, A., Garinella, K. et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1257–1303.

Meinke, D.W. (2020) Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. New Phytol. 226, 306–325.

Miten, R. (1992) Leaf glucosinolate profiles and their relationship to pest and disease resistance in oilseed rape. Euphytica, 63, 71–83.

Murphy, D.J. and Cummins, I. (1989) Biosynthesis of seed storage products during embryogenesis in rapeseed, Brassica napus. J. Plant Physiol. 135, 63–69.

Nord-Eilir, H.H., Andersen, T.G., Burow, M., Madsen, S.R., Jorgensen, M.E., Olsen, C.E., Dreyer, I. et al. (2012) NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. Nature, 488, 531–534.

Nord-Eilir, H.H., Madsen, S.R., Engelen, S., Jorgensen, M.E., Olsen, C.E., Andersen, J.S., Seynnaeve, D. et al. (2017) Reduction of antinutritional glucosinolates in Brassica oilseeds by mutation of genes encoding transporters. Nat. Biotechnol. 35, 377–382.

Okushima, Y., Mitina, I., Quach, H.L. and Theologis, A. (2005) AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. Plant J. 43, 29–46.

Parker, J.L. and Newstead, S. (2014) Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. Nature, 507, 68–72.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods, 14, 417–419.

Petersen, B.L., Chen, S.X., Hansen, C.H., Olsen, C.E. and Hawk, B.A. (2002) Composition and content of glucosinolates in developing Arabidopsis thaliana. Planta, 214, 562–571.

Redovnikovic, I.R., Tsev, S., Linsky, B. and Gershenson, J. (2012) Expression pattern of the glucosinolate side chain biosynthetic genes MAM1 and MAM3 of Arabidopsis thaliana in different organs and developmental stages. Plant Physiol. Biochm. 50, 77–83.

Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., Godde, M. et al. (2001) Bus, a bushy arabiadipos CYP79F1 knockout mutant with abolshed synthesis of short-chain aliphatic glucosinolates. Plant Cell, 13, 351–367.

Schneid, R. and Bencroft, I. (2011) Genetics and Genomics of the Brassicaceae. New York: Springer.

Schuff, M.C., Spielman, M., Tiwari, S., Adams, S., Fenby, N. and Scott, R.J. (2006) The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. Development, 133, 251–261.

Schuster, J., Knill, T., Reichelt, M., Gershenzon, J. and Binder, S. (2006) BRANCHED-CHAIN AMINOTRANSFERASE is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in Arabidopsis. Plant Cell, 18, 2664–2679.

Schweizer, F., Fernandez-Calvo, P., Zander, M., Diez-Diaz, M., Forseca, S., Glauser, G., Lewsey, M.G. et al. (2013) Arabidopsis basic helix-loop-helix
transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. Plant Cell Online, 25, 3117–3132.

Schwender, J., Shachar-Hill, Y. and Ohlrogge, J.B. (2006) Mitochondrial metabolism in developing embryos of Brassica napus. J. Biol. Chem. 281, 34040–34047.

Sharpe, A.G. and Lydiate, D.J. (2003) Mapping the mosaic of ancestral genotypes in a cultivar of oilseed rape (Brassica napus) selected via pedigree breeding. Genome, 46, 461–468.

Shi, L.L., Song, J.R., Guo, C.C., Wang, B., Guan, Z.L., Yang, P., Chen, X. et al. (2019) A CACTA-like transposable element in the upstream region of BnaA9.CYP79B9 acts as an enhancer to increase silique length and seed weight in rapeseed. Plant J. 98, 524–539.

Solcan, N., Kwock, J., Fowler, P.W., Cameron, A.D., Drew, D., Iwata, S. and Newstead, S. (2012) Alternating access mechanism in the POT family of oligopeptide transporters. Embo J. 31, 3411–3421.

Sonderby, I.E., Geu-Flores, F. and Halkier, B.A. (2010) Biosynthesis of glucosinolates—gene discovery and beyond. Trends Plant Sci. 15, 283–290.

Sonderby, I.E., Hansen, B.G., Bjørnholt, N., Ticcioni, C., Halkier, B.A. and Klebension, D.J. (2007) A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. Plos ONE, 2, e1322.

Soneson, C., Love, M.I. and Robinson, M.D. (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research, 4, 1521.

Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D. et al. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat. Plants, 6, 34–45.

Sotelo-Silveira, M., Cucinotta, M., Chauvin, A.L., Montes, R.A.C., Colombo, L., Marsch-Martinez, N. and de F Slater, S. (2013) Cytochrome F450 CYP79B9 is involved in Arabidopsis reproductive development. Plant Physiol. 172, 769–799.

Sun, J., Bankston, J.R., Payandeh, J., Hinds, T.R., Zagota, W.N. and Zheng, N. (2014) Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. Nature, 507, 73–77.

Tang, S., Zhao, H., Yu, L., Zhang, G., Zhang, Y., Yang, Q.Y. et al. (2019) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol. 14, 327.

Yu, B.Y., Gruber, M., Khachatourians, G.G., Hegedus, D.D. and Hannoufa, O. (2010) Gene expression profiling of developing Brassica napus seed in relation to changes in major storage compounds. Plant Sci. 178, 381–389.

Yu, H.J., Hogan, P. and Sundaresan, V. (2005) Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. Plant Physiol. 139, 1853–1869.

Ya, X.H., Gou, I.Y. and Liu, C.J. (2009) BAHD superfamily of acyl-CoA dependent acyltransferases in Populus and Arabidopsis: bioinformatics and gene expression. Plant Mol. Biol. 70, 421–442.

Zang, Y.X., Kim, J.H., Park, Y.D., Kim, D.H. and Hong, S.B. (2008) Metabolic engineering of aliphatic glucosinolates in Chinese cabbage plants expressing Arabidopsis CYP79F1 and CYP83A1. J. BMB Rep. 41, 472–478.

Zhang, B., Liu, C., Wang, Y., Yao, X. and Liu, K. (2015) Disruption of a CAROTENOID CLEAVAGE DIDYXYGENASE 4 gene converts flower colour from white to yellow in Brassica species. New Phytol. 206, 1513–1526.

Zhao, J. and Meng, J. (2003) Detection of loci controlling seed glucosinolate content and their association with Sclerotinia resistance in Brassica napus. Plant Breed. 122, 19–23.

Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824.

Zimmermann, I.M., Heim, M.A., Weisshaar, B. and Uhrig, J.F. (2010) Functional analysis of the tandem-duplicated P450 genes SPS/BUS/CYP79F1 and CYP79B2 in glucosinolate biosynthesis and plant development by Ds transposition-generated double mutants. Plant Physiol. 150, 840–848.

Supporting information

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

Figure S1 Histograms and correlation coefficients of BLUP values and the 3-year distribution of SGC in Wuhan for 505 accessions. Figure S2 GWAS of SGC for 2017–2019 in Wuhan. Figure S3 GO enrichment of genes significant in TWAS for SGC. Figure S4 Soft threshold power analysis to obtain scale-free fit metrics for network topology at 20 DAF (A) and 40 DAF (B). Figure S5 Prioritization of the causal genes in qGSL CO2.1. Figure S6 Analysis of nucleotide diversity (ø) and QTl haplotypes. Figure S7 In A–I, the boxplot of SGC between the two types of haplotypes of different QTLs in SW1 and SW2. Figure S8 Prioritization of the causal genes in qGSL A09.2. Figure S9 Prioritization of the causal genes in qGSL CO9.1. Figure S10 Boxplots for TSW and SOC. Figure S11 Characterization of editing of BnaGTR2 genes using the C1RSPR/Cas9 system. Figure S12 Comparison of the diameter of mature seeds between Bna.gtr2 mutants. Figure S13 Boxplot of seed traits in WT and Bna.gtr2 mutants. Figure S14 GSL content in leaves (A) and roots (B) (µmol/g) of WT and BnaC02.gtr2 mutants.
Figure S15 Seed development-related genes overlapped with genes significant in TWAS in SGC.

Dataset S1 Dataset S2 Dataset S3 Dataset S4 Dataset S5 Dataset S6 Dataset S7 Dataset S8 Dataset S9 Dataset S10 Dataset S11 Dataset S12 Dataset S13 Dataset S14 Dataset S15 Dataset S16 Dataset S17 Dataset S18 Dataset S19 Dataset S20 Dataset S21 Dataset S22 Dataset S23 Dataset S24 Dataset S25 Dataset S26 Dataset S27 Dataset S28 Dataset S29 Dataset S30

Dataset S1 Dataset S2 Dataset S3 Dataset S4 Dataset S5 Dataset S6 Dataset S7 Dataset S8 Dataset S9 Dataset S10 Dataset S11 Dataset S12 Dataset S13 Dataset S14 Dataset S15 Dataset S16 Dataset S17 Dataset S18 Dataset S19 Dataset S20 Dataset S21 Dataset S22 Dataset S23 Dataset S24 Dataset S25 Dataset S26 Dataset S27 Dataset S28 Dataset S29 Dataset S30

ª 2021 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 20, 211–225

Genetic basis of seed glucosinolate content in rapeseed