Genome-wide association study identifies new loci for 1000-seed weight in Brassica napus

Haijiang Liu
Jingchi Wang
Bingbing Zhang
Xinyu Yang
Pan Yuan
Guangda Ding
Sheliang Wang
Hongmei Cai
Chuang Wang
Fangsen Xu
LEI SHI (leish@mail.hzau.edu.cn)
National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University
https://orcid.org/0000-0002-5312-8521

Research Article

Keywords: Oilseed rape, genome wide association study, 1000-seed weight, haplotype analysis

Posted Date: April 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1527829/v1

Abstract

Oilseed rape (*Brassica napus* L.) is an important oilseed crop worldwide and 1000-seed weight (SW) is the important determinant of seed yield of *B. napus*. To elucidate the genetic mechanism of SW and mining candidate genes, a panel of 403 diverse *B. napus* accessions was screened in a genome-wide association study (GWAS) using 5.58 million single-nucleotide polymorphisms (SNPs). This study identified 340 SNPs significantly associated with SW by general linear model (GLM) and mixed linear model (MLM). Through GWAS combined with transcriptome data, two significantly differentially expressed genes were identified as candidate genes (*BnaA02g06870D* and *BnaC06g28920D*). Candidate gene association analysis and haplotype analysis showed that the inbred lines carrying ACCC at *BnaA02g06870Hap1* and TTGG at *BnaC06g28920Hap1* had greater SW than lines carrying other haplotype alleles. Candidate genes and favourable haplotypes identified in this study will be useful for large-seed breeding of *B. napus*.

Introduction

Oilseed rape (*Brassica napus* L.) is an important oil crop that is used as edible oil and a renewable energy resource. In 2020, global vegetable oil production has reached 75 million tons (http://www.fao.org/faostat/zh/#data/QC/visualize). Seed yield of *B. napus* is determined by 1000-seed weight (SW), siliques per plant and seed number per silique (Fan et al. 2010; Lu et al. 2017a). Several genes that control SW have been identified, such as *Atsob3-6* and *AtENO2* in *Arabidopsis*; *ZmBES1* in maize; *OsMED15a*, *OsMKKK10*, *OsMKK4*, and *OsMAPK6* in rice; *GmCYP78A72* and *GmBZR1* in soybean; and *BnARF18* in oilseed rape (Liu et al. 2015; Zhao et al. 2016a; Lu et al. 2017b; Xu et al. 2018; Dwivedi et al. 2019; Liu et al. 2020; Sharma et al. *20*20; Sun et al. 2021). However, most genes associated with SW have not been excavated due to its complexity of genetic mechanism (Wang et al. 2020).

Linkage mapping analysis has been widely used to study the genetic basis of SW in *B. napus* (Shi et al. 2009; Zhao et al. 2016b; Wang et al. 2020). One hundred and fifty-nine quantitative trait loci (QTLs) are significantly associated with SW in *B. napus* on the analysis of two (TNDH and RC-F2) populations in ten natural environments, and one major QTL was detected in ten environments (Shi et al. 2009). In addition, twenty-one significant QTLs are associated with seed yield and yield-related traits from three different trials (Zhao et al. 2016b). Among them, nine QTLs were associated with SW, explaining 4.58–19.62% of the phenotypic variation (PVE) (Zhao et al. 2016b). Recently, a significant QTL (*cqSW.A03-2*) associated with SW is detected across multiple environments, explaining 8.46–13.70% of the PVE (Wang et al. 2020). The subsequent candidate gene association analysis and gene expression analysis show that *BnaA03g37960D* is the most likely candidate gene for the significant QTL (Wang et al. 2020).

The physical intervals for QTLs are usually very large, which makes it difficult to identify candidate genes (Xiao et al. 2017). So far, only two genes (*BnaA09.ARF18* and *BnaA9.CYP78A9*) controlling SW have been cloned in *B. napus* using linkage analysis (Liu et al. 2015; Shi et al. 2019). Compared with QTL mapping, GWAS take advantage of PVE and historical recombination in natural populations, which greatly

improves the efficiency of gene mapping (Nordborg and Weigel 2008). GWAS has become a common method to analyze the genetic structure of complex agronomic traits, and it has also been widely used in mining genes associated with seed (grain) weight, such as in rice (Tao et al. 2019; Niu et al. 2021), maize (Zhang et al. 2020), soybean (Qi et al. 2020; Zhang et al. 2021) and oilseed rape (Li et al. 2014a; Lu et al. 2017a). With the maturity of technology and the decline of sequencing price, researchers have developed SNP markers through whole-genome resequencing technology in *B. napus* (Wu et al. 2019; Lu et al. 2019; Tang et al. 2021). Some traits previously analyzed by GWAS using 60K SNP chip as genotype have been re-analyzed, and some new loci and candidate genes have been excavated, such as, flowering time (Zhou et al. 2017), glucosinolate content (Wei et al. 2019; Tan et al. 2022) and seed oil content (Xiao et al. 2019; Tang et al. 2021; Xiao et al. 2021). In this study, a diverse panel of 403 *B. napus* accessions were scored for SW and 340 significant SNPs were identified by GWAS. Two genes, *BnaA02g06870D* and *BnaC06g28920D*, were identified as candidate genes, and the favourable haplotypes (*BnaA02g06870Hap1* and *BnaC06g28920Hap1*) were revealed for breeding large-seed *B. napus* cultivars. These results will contribute to improve our understanding of the genetic mechanism of SW of *B. napus* and breeding of large-seed *B. napus* cultivars.

Materials And Methods

Plant materials and field trails

Detailed information on the 403 diverse *B. napus* accessions and genotypes are described in our previous study (Liu et al. 2021). All *B. napus* accessions were grown in the field with three replications at Meichuan Town, Wuxue city, Hubei province, China (115·55° E , 29·85° N) from 2018 to 2019 (Trial 1) and from 2019 to 2020 (Trial 2). Each accession was grown in a plot with four rows, and each plot had eight plants in each row, with a distance of 25 cm between plants in each row and 25 cm between rows. After harvest, six plants for each repeat of each accession were selected to investigate SW.

Genome-wide association analysis and candidate gene identification

More than 10 million high-quality SNPs of the association panel were derived from a previous study and after filtering the SNPs with minor-allele frequency (MAF) >0.05 and missing rate <0.2, a total of 5.58 million SNPs were used for GWAS (Tang et al. 2021; Liu et al. 2021). Genome-wide association analysis for SW was carried out using general linear models (GLMs) and mixed linear models (MLMs) by Tassel 5.0 software (Bradbury et al. 2007). The significant P-value thresholds for the association panel was 6.25×10^{-07}. GGplot2 (https://cran.r-project.org/web/packages/ggplot2/index.html) software was used to draw Manhattan plot, and CMplot software was used to draw Quantile–Quantile plot (https://github.com/YinLiLin/CMplot). The LD decay, population structure and kinship of this association panel have been reported in previous studies (Liu et al. 2021). The genes located within LD decay value upstream and downstream of the peak SNPs were considered as candidate genes. The genotypes of *BnaA02g06870D* and *BnaC06g28920D* in the association panel of *B. napus* were obtained by vcftools software (https://vcftools.github.io/index.html). Candidate gene association analysis of

BnaA02g06870D and *BnaC06g28920D* were performed with Tassel 5.0 software (Bradbury et al. 2007). The SNP markers from 2 kb upstream of the gene to termination codon were used to conduct association analysis with the SW of the association panel of *B. napus*.

Haplotype analysis

HaploView.4.2 software was used to conduct haplotype analysis (Barrett et al. 2005). Haplotypes containing at least 10 *B. napus* accessions were used for further comparative analysis, and Student's t-test was used to compare the differences in SW among the haplotypes.

2.4 Statistical analysis of phenotypic data

The mean value, maximum, minimum and coefficient of variation were calculated using Excel 2007. The R language was used to calculate the correlation coefficients between trials. The broad-sense heritability was calculated as: $h^2 = V_G / (V_G / (V_G + V_E) / nr)$, where V_G is genetic variance, V_E is environmental variance, n is the number of environments and r is the number of replicates.

Results

Phenotypic variation for SW of an association panel of B. napus

SW was investigated for the association panel of 403 *B. napus* accessions in 2018–19 (Trial 1) and 2019–20 (Trial 2) field trials. Extensive phenotypic variations for SW were observed in the association panel of *B. napus* (Fig. S1; Table 1, Table S1). For example, SW ranged from 3.30 to 4.60 g/ 1000 seeds (1.4-fold) in Trial 1 and from 3.00 to 5.08 g/ 1000 seeds (1.69-fold) in Trial 2 (Table 1). In addition, high h^2 values were observed for SW (Table 1). The correlation coefficient (r) of SW between Trial 1 and Trial 2 were 0.64, which showed that the phenotype had good repeatability in two trials (Fig. S2).

Table 1
Mean, maximum (max), minimum (min), coefficient of variation (CV, %) and heritability (h^2) of 1000-seed weight (SW) in an association panel of *B. napus* in Trial 1 and Trial 2

Year	Mean	Min	Max	CV	h^2 (%)
Trial 1	3.46	3.30	4.64	16.57%	78.64%
Trial 2	4.04	3.00	5.08	12.55%	

Genome-wide association study of SW of B. napus

We performed GWAS with GLM and MLM approaches to identify SNPs associated with SW in *B. napus*. A total of 340 SNPs were significantly associated with SW across two trials ($P < 6.25 \times 10^{-07}$) (Fig. 1; Table S2; Table S3). Among the 340 SNPs, 180 were identified in Trial 1, 16 in Trial 2, and 144 were identified by

the mean values of two trials (Fig. 1; Table S2; Table S3). The GLM analysis detected a total of 340 SNPs significantly associated with SW, distributed on 18 of the 19 *B. napus* chromosomes (excluding A08). Chromosomes C06 and A02 had the largest number of significant SNPs (32) and the second largest number of significant SNPs (61 SNPs), respectively (Table S2). Since MLM considers both population structure and kinship, only 3 SNPs were detected by MLM, which distributed on chromosomes A02, C06 and C07 of *B. napus*, explaining the PVE of 9.72–10.25% (Fig. 1D, E, F; Table S3). These three SNPs were detected simultaneously by GLM and MLM (Table S2; Table S3). Additionally, 10.30% (35/340) of the significant SNPs were identified in more than one trial (including the mean value of two trials), which showed high reliability (Fig. 1, Table S2). For example, the significant SNP marker (chrA02__3219827) was detected by Trial 1, Trial 2 and the mean value of two trials, which could explain 12.81% of the PVE of SW (Fig. 1A, B, C; Table S2). The significant SNP marker (chrC06__29452686) was detected by Trial 1, Trial 2 and the mean value of two trials, which could explain 9.28% of the PVE of SW (Fig. 1A, B, C; Table S2).

Comparison of significant SNPs by GWAS with reported QTLs for SW in B. napus

Based on the Darmor-*bzh* reference genome, we analysed the co-localization of the significant SNPs detected in our study and the QTLs detected by previous studies (Basunanda et al. 2010; Yang et al. 2012; Li et al. 2014b; Fu et al. 2015; Liu et al. 2015; Wang et al. 2020). Nine significant SNPs detected by GWAS co-localized with the intervals of the QTLs for SW in a previous linkage analyses, including seven SNPs on A03, and two SNPs on A09 (Table 2). For example, the SNP ‘chrA03__18945635’ on chromosome A03 was detected for SW in Trial 1, and explained 8.83% of the PVE, which co-located with the reported QTL (*cqSW.A03-2*) (Table 2). The SNP ‘chrA09__27535146’ on chromosome A09 was detected for SW with the mean value of the two trials, and explained 9.34% of the PVE (Table 2), which co-located with the reported QTL (*cqSW.A09-3*) (Table 2).

Table 2
Co-location of SNPs associated with 1000-seed weight (SW) detected by GWAS with an association panel of *B. napus* and QTLs for SW detected in previous studies.

SNP	Chr	Pos. (bp)	P value	PVE (%)	Experiment	References (Previous QTL loci)
chrA03__18945635	A03	18945635	9.56E-08	8.83%	Trial 1	Wang et al. 2020
chrA03__19955110	A03	19955110	3.24E-07	8.53%	Trial 1	
chrA03__19955115	A03	19955115	2.09E-07	8.67%	Trial 1	
chrA03__19955118	A03	19955118	1.76E-07	8.98%	Trial 1	
chrA03__18022289	A03	18022289	6.37E-07	7.48%	The mean value of two trials	
chrA03__18034286	A03	18034286	2.02E-07	8.90%	The mean value of two trials	
chrA03__18091274	A03	18091274	4.28E-07	7.39%	The mean value of two trials	
chrA09__27535146	A09	27535146	2.26E-07	8.76%	The mean value of two trials	Basunanda et al. 2010; Yang et al. 2012; Li et al. 2014; Fu et al. 2015; Liu et al. 2015
chrA09__27915575	A09	27915575	1.16E-07	9.34%	The mean value of two trials	

New loci and candidate genes for SW in B. napus

The SNP of chrA02__3219827 on chromosome A02 was associated with SW in both Trial 1, Trial 2 and the mean value of two trials (Fig. 1A, B, C; Table S2). The LD decay of chromosome A02 in this association panel was 174 kb (Liu et al. 2021). Based on the LD decay, 174 kb up/down-stream of the significant SNP (chrA02__3219827) was selected to identify candidate genes on A02 and 82 candidate genes were detected (Table S4). Previously, a transcriptome analysis of the immature seeds of two *B. napus* lines with extremely different SW has been conducted to identify the candidate genes related to SW (Geng et al. 2018). In this study, a comparison of our GWAS and transcriptome sequencing results by Geng et al. (2018) revealed 2 common genes (*BnaA02g06330D* and *BnaA02g06870D*) on A02 chromosome (Table 3). In the seeds of large-seed line, the expression levels of *BnaA02g06330D* and *BnaA02g06870D* were 4.18 and 3.02 fold higher than those of small-seed line, respectively (Table 3). Twelve SNPs were located within the 2 kb promoter region and the entire coding region of *BnaA02g06870D* (Table S4; Table S5). However, no SNP was identified within the corresponding region of

BnaA02g06330D (Table 3). We performed candidate gene association analysis of *BnaA02g06870D*, and four SNPs in *BnaA02g06870D* were significantly associated with SW in Trial 1, Trial 2 and the mean value of two trials (Fig. 2A, B, C). Further analysis demonstrated that the A allele of chrA02__3257485, C allele of chrA02__3257503, C allele of chrA02__3257506 and C allele of chrA02__3257518 were the large-seed alleles (Fig. 3). Two major haplotypes were detected, and cultivars with *BnaA02g06870Hap1* (ACCC) had much higher SW than cultivars with *BnaA02g06870Hap2* (TTTT), in Trial 1, Trial 2 and the mean value of two trials. Thus *BnaA02g06870Hap1* was confirmed as the favorable haplotype (Fig. 2D, E, F).

Table 3
Common candidate genes associated with 1000-seed weight identified by GWAS and transcriptome sequencing

Gene ID	Log_2(fold change)	P value	Number of SNPs in genes	Arabidopsis homologue	Function description
BnaA02g06330D	4.18	1.28E-07	0	*AT5G58470.2*	TBP-associated factor 15B (TAF15b); FUNCTIONS IN: binding, nucleotide binding, zinc ion binding, nucleic acid binding
BnaA02g06870D	3.02	1.14E-07	12	*AT5G59740.1*	UDP-N-acetylglucosamine (UAA) transporter family; FUNCTIONS IN: galactose transmembrane transporter activity
BnaC06g28920D	3.04	1.11E-06	20	*AT1G17770.1*	SU(VAR)3–9 homolog 7 (SUVH7); FUNCTIONS IN: zinc ion binding, histone-lysine N-methyltransferase activity

The confidence region significantly associated with SW in both Trial 1, Trial 2 and the mean value of two trials, on chromosome C06 ranged from 29.47 to 31.25 Mb (Fig. 1A, B, C; Table S2). The lead SNP ‘chrC06__30203895’ on chromosome C06 was detected for SW in Trial 1, and explained 11.30% of the PVE (Table S2). In previous study, the LD decay of C06 chromosome in this association panel was 229 kb (Liu et al. 2021). Based on the LD decay, 229 kb upstream and downstream regions of the significant SNP (chrC06__30203895) were selected and found to contain 81 genes (Table S6). One significantly differentially expressed gene (*BnaC06g28920D*) was detected by combining GWAS and the transcriptome data by Geng et al. (2018) (Table 3). The expression levels of *BnaC06g28920D* in the seeds of large-seed line were 3.04 fold higher than that of small-seed line (Table 3). Twenty SNPs located within the 2 kb promoter region and the entire coding region of *BnaC06g28920D* (Table 4). Candidate gene association analysis showed that, five SNPs in *BnaC06g28920D* were detected to be significantly associated with the SW in Trial 1, six in Trial 2 and seven with the mean value of two trials (Fig. 4A, B, C). Notably, chrC06__30094959 (T/C) were located in the exon region of the gene *BnaC06g28920D* and resulted in amino acid changes from isoleucine to threonine; chrC06__30095553 (T/C) were located in the exon

region of the gene *BnaC06g28920D* and resulted in amino acid changes from leucine to proline; chrC06__30095570 (G/T) were located in the exon region of the gene *BnaC06g28920D* and resulted in amino acid changes from valine to leucine; and chrC06__30096038 (G/C) were located in the exon region of the gene *BnaC06g28920D* and resulted in amino acid changes from valine to leucine (Table 4). We observed that the T allele of chrC06__30094959, T allele of chrC06__30095553, G allele of chrC06__30095570, and G allele of chrC06__30096038 were large-seed alleles (Fig. 5). These four significant SNPs revealed four haplotypes, and *BnaC06g28920Hap1* (TTGG) had significantly greater SW value than other haplotypes in Trial 1, Trial 2 and the mean value of two trials. *BnaC06g28920Hap1* was a favorable haplotype (Fig. 4D, E, F). In addition, a total of 44 *B. napus* cultivars contained these eight favorable alleles, and they had higher SW than *B. napus* cultivars with other alleles (Fig. 6).

Table 4
List of synonymous and non-synonymous SNP variants identified in the candidate gene of *BnaC06g28920D*

SNP	Major allele	Minor allele	SNP location	SNP types	Amino acid changes
chrC06__30094868	C	T	Exon	Non-synonymou	Proline to serine
chrC06__30094959	T	C	Exon	Non-synonymou	Isoleucine to threonine
chrC06__30095014	A	G	Exon	Synonymous	-
chrC06__30095467	T	C	Exon	Synonymous	-
chrC06__30095553	T	C	Exon	Non-synonymous	Leucine to proline
chrC06__30095570	G	T	Exon	Non-synonymous	Valine to leucine
chrC06__30095585	T	C	Exon	Synonymous	-
chrC06__30095596	T	C	Exon	Synonymous	-
chrC06__30095632	G	A	Exon	Synonymous	-
chrC06__30095668	C	T	Exon	Synonymous	-
chrC06__30095711	A	G	Exon	Synonymous	-
chrC06__30095762	A	C	Exon	Non-synonymous	Lysine to glutamine
chrC06__30095770	C	T	Exon	Synonymous	-
chrC06__30095926	A	G	Exon	Synonymous	-
chrC06__30095928	A	T	Exon	Non-synonymous	Tyrosine to Phenylalanine
chrC06__30095944	A	G	Exon	Synonymous	-
chrC06__30096004	G	A	Exon	Synonymous	-
chrC06__30096013	T	C	Exon	Synonymous	-
chrC06__30096037	G	A	Exon	Synonymous	-
chrC06__30096038	G	C	Exon	Non-synonymous	Valine to leucine

Note: SNP, Single-nucleotide polymorphism.

Discussion

Overlapped loci for SW between this study and previous studies

SW is an important yield related trait and breeding large seed *B. napus* is a long-term goal of breeders (Fan et al. 2010; Wang et al. 2020). Mining genes controlling SW and finding excellent haplotypes associated with SW can promote molecular marker assisted breeding of *B. napus*. A panel of 472 inbred lines of *B. napus* were genotyped using 24256 SNPs, and two SNPs were associated with SW, which distributed on A07 and A09 chromosomes, and explained 4.90% and 13.87% of the PVE, respectively (Li et al. 2014a). In another study, nine SNPs were significantly associated with SW on A02, A03, A08, A09, A10, C03, C04 and C09 chromosomes by using a panel of 520 *B. napus* accessions and 31,839 SNPs (Lu et al. 2017a). In addition, seven SNPs were significantly associated with SW with 157 *B. napus* cultivars and 690953 SNPs, (Dong et al. 2018). In our study, a panel of 403 diverse *B. napus* accessions was screened in a GWAS using more than five million SNPs and identified 340 significant SNPs associated with SW (Table S2; Table S3). Twenty-two significant SNPs in this study were adjacent to previously published significant loci for SW, which increased the credibility of these SNPs (Table S7). For example, chrA09__33133565 explained 7.78% of PVE of SW, which was adjacent to the significant SNP (Bn-A09-p30654305) located in Li et al. (2014a) (Table S7). Moreover, chrA02__3573374, a lead SNP on the A02 chromosome was detected by SW in Trial 1, which was adjacent to the significant SNP (rs4600) located in Lu et al. (2017a) (Table S7). In addition to GWAS method, some QTLs related to SW were also mined by linkage mapping. The intervals of a major QTL controlling SW of *B. napus* were reported overlapped in different studies (Basunanda et al. 2010; Yang et al. 2012; Li et al. 2014b; Fu et al. 2015; Liu et al. 2015; Dong et al. 2018) (Table 2; Table S7). Subsequent studies show that *BnaARF18* in this interval regulates cell growth in the silique wall by acting via an auxin-response pathway and finally changed the SW (Liu et al. 2015). In this study, SNP chrA09__27915575 explained 9.34% of PVE of SW, which was co-located with the significant QTL of *cqSW.A09* (Table 2; Table S7).

In addition, the lead SNPs of chrA02__3219827 and chrC06__29452686 were both associated with SW in Trial 1, Trial 2 and the mean value of two trials (Fig. 1A, B, C; Table S2). These SNPs showed higher reliability, and they were newly found important loci for SW, which were not detected in previous studies.

Causal genes associated with the significant loci for SW

The LD decay values in several GWAS studies on SW of *B. napus* were 1.2 Mb (472 inbred lines and 24256 SNPs), 0.8 Mb (520 inbred lines and 31839 SNPs) and 0.6 Mb (520 inbred lines and 690953 SNPs), respectively (Li et al. 2014a; Lu et al. 2017a; Dong et al. 2018). Large LD decay in *B. napus* indicates an associated locus contains more genes, and thus it is difficult to pinpoint the causal genes associated with the significant loci. Although the LD decay (237kb) value in this association panel is much smaller than that in previous studies, however, there are still dozens of genes in the interval, so it is also difficult to determine the real candidate genes. For example, there were 82 candidate genes within the range of LD decay value upstream and downstream the lead significant SNP (chrA02__3219827) on the A02 chromosome (Table S4).

GWAS combined with transcriptome analysis has become a common combinatorial analysis for mining candidate genes in *B. napus* in recent years (Lu et al. 2016; Wei et al. 2016; Xiao et al. 2019). For example, Xiao et al (2019) combined GWAS and transcriptome to mine seed oil content-related candidate genes and identified seven candidate genes (Xiao et al. 2019). In addition, a GWAS combined with transcriptome identified 33 significantly differentially expressed genes candidate genes located within the confidence intervals of significant SNPs for harvest index (Lu et al. 2016). In this study, the reported transcriptome data of the immature seeds of two *B. napus* lines with extremely different SW (Geng et al. 2018) were employed, and *BnaA02g06870D* and *BnaC06g28920D* were predicted to be candidate genes.

Favorable alleles, haplotypes and large-seed breeding

Candidate gene association study can mine the genetic variation significantly related to phenotype in candidate genes (Ma et al. 2021; Zheng et al. 2017; Yu et al. 2019; Wang et al. 2020). For example, the favorable alleles of candidate gene '*BnaA03g37960D'* for SW were identified in *B. napus* and SW of cultivars with two different SNPs (SA03_18861910 and SA03_18862302) are significantly different (Wang et al. 2020). In another study, the significant SNP Bn-A02-p9505646 (G/A) of the GG allele showed the largest contribution (R^2 ~9.79%) to plant height of *B. napus*, and the plant height of plants with GG allele are far higher than those with AA allele. The significant SNP Bn-A02-p9505646 (C/A) of the CC allele showed the largest contribution (R^2 ~9.87%) to branch initation height regulation, measuring 11.4 cm more than those of the AA allele (Zheng et al. 2017). In this study, we performed candidate gene association analysis on two candidate genes (*BnaA02g06870D* and *BnaC06g28920D*) and mined a total of eight significant SNPs (Fig. 2; Fig. 4). The SW of cultivars with favorable alleles was higher than that of other cultivars (Fig. 3; Fig. 5; Fig. 6).

The genetic variation in promoter and 5'UTR region may cause the difference of candidate gene expression among cultivars and the genetic variation in exon region may cause the change of amino acids among cultivars, which can affect the function of the candidate gene protein (Farashi et al. 2019). In this study, four SNPs (chrA02__3257485, chrA02__3257503, chrA02__3257506 and chrA02__3257518) significantly related to seed weight were detected in candidate gene *BnaA02g06870D*, which were located in the promoter region of *BnaA02g06870D* (Fig. 2; Table S5). In another candidate gene *BnaC06g28920D*, four SNP loci (chrC06__30094959, chrC06__30095553, chrC06__30095570 and chrC06__30096038) were significantly related to SW (Fig. 4; Table 4). These SNPs were located in the exon region and caused amino acid changes among cultivars (Fig. 4; Table 4). Subsequently, the favorable haplotypes of *BnaA02g06870D* and *BnaC06g28920D* were determined as ACCC and TTGG by haplotype analysis, respectively. The cultivars carrying ACCC at *BnaA02g06870Hap1* and TTGG at *BnaC06g28920Hap1* had greater SW than other cultivars (Fig. 2D, E, F; Fig. 4D, E, F). The favorable haplotypes ACCC and TTGG identified in this study can be used for molecular marker-assisted breeding of large seed in *B. napus*.

In conclusion, SW of *B. napus* is a complex quantitative trait controlled by multiple genes. A total of 340 SNPs were identified to be significantly associated with SW across two years, and thirty-five of them were detected simultaneously for different years. Two new candidate genes for SW were identified

(*BnaA02g06870D* and *BnaC06g28920D*) and the discovery of the large seed haplotype of ACCC at *BnaA02g06870Hap1*, and TTGG at *BnaC06g28920Hap1* could enable the accurate selection of *B. napus* cultivars with large seed.

Abbreviations

SW 1000-seed weight; GWAS genome-wide association study; SNP single-nucleotide polymorphisms; GLM general linear model; MLM mixed linear model; QTL quantitative trait loci; PVE phenotypic variation;

Declarations

Acknowledgements

This work was supported by the National Nature Science Foundation of China (Grant Nos. 32172662 and 31972498). We also acknowledge the Applied Basic Research Fronts Program of Wuhan city (Grant No. 2018020401011302), the National Key R &D Program of China (Grant No. 2017YFD0200200) and the Natural and Fundamental Research Funds for the Central Universities of China (Grant No. 2662019PY013). The computations in this paper were run on the bioinformatics computing platform of the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University.

Author contribution statement

Haijiang Liu and Lei Shi designed research, reviewed writing and drafted the manuscript. Haijiang Liu, Jingchi Wang, Bingbing Zhang, Xinyu Yang and Pan Yuan: participated the experiments. Lei Shi,Guangda Ding, Sheliang Wang, Hongmei Cai, Chuang Wang and Fangsen Xu: Participated in manuscript revision

Data Availability

The association mapping population data can be found the Genome Sequence Archive (https://bigd.big.ac.cn/gsa/) with Bioproject IDs PRJCA002835 and PRJCA002836 (Tang et al. 2021), and transcriptome data can be found in the Gene Expression Omnibus (GEO) database, with the accession number SRR9165867 and SRR9165099 (Geng et al. 2018).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Human and animal rights

This study does not include human or animal subjects.

References

1. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263–265. https://doi.org/10.1093/bioinformatics/bth457
2. Basunanda P, Radoev M, Ecke W, Friedt W, Becker HC, Snowdon RJ (2010) Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). Theor Appl Genet 120(2):271–281. https://doi.org/10.1007/s00122-009-1133-z
3. 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(19):2633–2635. https://doi.org/10.1093/bioinformatics/btm308
4. Dong H, Tan C, Li Y, He Y, Wei S, Cui Y, Chen Y, Wei D, Fu Y, He Y, Wan H, Liu Z, Xiong Q, Lu K, Li J, Qian W (2018) Genome-wide association study reveals both overlapping and independent genetic loci to control seed weight and silique length in *Brassica napus*. Front Plant Sci 9:921. https://doi.org/10.3389/fpls.2018.00921
5. Dwivedi N, Maji S, Waseem M, Thakur P, Kumar V, Parida SK, Thakur JK (2019) The Mediator subunit *OsMED15a* is a transcriptional co-regulator of seed size/weight-modulating genes in rice. Biochim Biophys Acta Gene Regul Mech 1862(10):194432. https://doi.org/10.1016/j.bbagrm.2019.194432
6. Fan C, Cai G, Qin J, Li Q, Yang M, Wu J, Fu T, Liu K, Zhou Y (2010) Mapping of quantitative trait loci and development of allele-specific markers for seed weight in *Brassica napus*. Theor Appl Genet 121(7):1289–1301. https://doi.org/10.1007/s00122-010-1388-4
7. Farashi S, Kryza T, Clements J, Batra J (2019) Post-GWAS in prostate cancer: from genetic association to biological contribution. Nat Rev Cancer 19(1):46–59. https://doi.org/10.1038/s41568-018-0087-3
8. Fu Y, Wei D, Dong H, He Y, Cui Y, Mei J, Wan H, Li J, Snowdon R, Friedt W, Li X, Qian W (2015) Comparative quantitative trait loci for silique length and seed weight in *Brassica napus*. Sci Rep 23(5):14407. https://doi.org/10.1038/srep14407
9. Geng X, Dong N, Wang Y, Li G, Wang L, Guo X, Li J, Wen Z, Wei W (2018) RNA-seq transcriptome analysis of the immature seeds of two *Brassica napus* lines with extremely different thousand-seed weight to identify the candidate genes related to seed weight. PLoS One 13(1):e0191297. https://doi.org/10.1371/journal.pone.0218914
10. Li F, Chen B, Xu K, Wu J, Song W, Bancroft I, Harper AL, Trick M, Liu S, Gao G, Wang N, Yan G, Qiao J, Li J, Li H, Xiao X, Zhang T, Wu X (2014a) Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). DNA Res 21(4):355–367. https://doi.org/10.1093/dnares/dsu002
11. Li N, Shi J, Wang X, Liu G, Wang H (2014b) A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (*Brassica napus* L.). BMC Plant Biol 14:114. https://doi.org/10.1186/1471-2229-14-114
12. Liu J, Hua W, Hu Z, Yang H, Zhang L, Li R, Deng L, Sun X, Wang X, Wang H (2015) Natural variation in *ARF18* gene simultaneously affects seed weight and silique length in polyploid rapeseed. Proc Natl Acad Sci U S A 112(37):E5123–5132. https://doi.org/10.1073/pnas.1502160112

13. Liu Z, Zheng L, Pu L, Ma X, Wang X, Wu Y, Ming H, Wang Q, Zhang G (2020) ENO2 Affects the seed size and weight by adjusting cytokinin content and forming ENO2-bZIP75 complex in *Arabidopsis thaliana*. Front Plant Sci 11:574316. https://doi.org/10.3389/fpls.2020.574316
14. Liu H, Wang J, Zhang B, Yang X, Hammond JP, Ding G, Wang S, Cai H, Wang C, Xu F, Shi L (2021) Genome wide association study dissects the genetic control of plant height and branch number in response to lowphosphorus stress in *Brassica napus*. Ann Bot 128(7):919–929. https://doi.org/10.1093/aob/mcab115
15. Lu K, Xiao Z, Jian H, Peng L, Qu C, Fu M, He B, Tie L, Liang Y, Xu X, Li J (2016) A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in *Brassica napus*. Sci Rep 6:36452. https://doi.org/10.1038/srep36452
16. Lu K, Peng L, Zhang C, Lu J, Yang B, Xiao Z, Liang Y, Xu X, Qu C, Zhang K, Liu L, Zhu Q, Fu M, Yuan X, Li J (2017a) Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in *Brassica napus*. Front Plant Sci 8:206. https://doi.org/10.3389/fpls.2017.00206
17. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H, Cheng F, Zhang K, Du H, Cheng X, Qu C, Qian W, Liu L, Wang R, Zou Q, Ying J, Xu X, Mei J, Liang Y, Chai YR, Tang Z, Wan H, Ni Y, He Y, Lin N, Fan Y, Sun W, Li NN, Zhou G, Zheng H, Wang X, Paterson AH, Li J (2019) Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat Commun 10(1):1154. https://doi.org/10.1038/s41467-019-09134-9
18. Lu X, Xiong Q, Cheng T, Li QT, Liu XL, Bi YD, Li W, Zhang WK, Ma B, Lai YC, Du WG, Man WQ, Chen SY, Zhang JS (2017b) A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. Mol Plant 10(5):670–684. https://doi.org/10.1016/j.molp.2017.03.006
19. Ma L, Zhang M, Chen J, Qing C, He S, Zou C, Yuan G, Yang C, Peng H, Pan G, Lubberstedt T, Shen Y (2021) GWAS and WGCNA uncover hub genes controlling salt tolerance in maize (Zea mays L.) seedlings. Theor Appl Genet 134(10):3305–3318. https://doi.org/10.1007/s00122-021-03897-w
20. Niu Y, Chen T, Wang C, Chen K, Shen C, Chen H, Zhu S, Wu Z, Zheng T, Zhang F, Xu J (2021) Identification and allele mining of new candidate genes underlying rice grain weight and grain shape by genome-wide association study. BMC Genomics 22(1):602. https://doi.org/10.1186/s12864-021-07901-x
21. Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720–723. https://doi.org/10.1038/nature07629
22. Qi Z, Song J, Zhang K, Liu S, Tian X, Wang Y, Fang Y, Li X, Wang J, Yang C, Jiang S, Sun X, Tian Z, Li W, Ning H (2020) Identification of QTNs controlling 100-Seed weight in soybean using multilocus genome-wide association studies. Front Genet 11:689. https://doi.org/10.3389/fgene.2020.00689
23. Sharma Koirala P, Neff MM (2020) Improving seed size, seed weight and seedling emergence in Camelina sativa by overexpressing the Atsob3-6 gene variant. Transgenic Res 29(4):409–418. https://doi.org/10.1007/s11248-020-00208-9

24. Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. Genetics 182(3):851–861. https://doi.org/10.1534/genetics.109.101642
25. Shi L, Song J, Guo C, Wang B, Guan Z, Yang P, Chen X, Zhang Q, King GJ, Wang J, Liu K (2019) A CACTA-like transposable element in the upstream region of *BnaA9.CYP78A9* acts as an enhancer to increase silique length and seed weight in rapeseed. Plant J 98(3):524–539. https://doi.org/10.1111/tpj.14236
26. Sun F, Ding L, Feng W, Cao Y, Lu F, Yang Q, Li W, Lu Y, Shabek N, Fu F, Yu H (2021) Maize transcription factor *ZmBES1/BZR1-5* positively regulates kernel size. J Exp Bot 72(5):1714–1726. https://doi.org/10.1093/jxb/eraa544
27. Tan Z, Xie Z, Dai L, Zhang Y, Hu Z, 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(1):211–225. https://doi/10.1111/pbi.13707
28. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang QY, Zhou Y, Wang X, Ma W, Xie W, 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*. Mol Plant 14(3):470–487. https://doi.org/10.1016/j.molp.2020.12.003
29. Tao Y, Zhao X, Wang X, Hathorn A, Hunt C, Cruickshank AW, Oosterom EJ, Godwin ID, Mace ES, Jordan DR (2019) Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. Plant Biotechnol J 18(4):1093–1105. https://doi.org/10.1111/pbi.13284
30. Wang H, Yan M, Xiong M, Wang P, Liu Y, Xin Q, Wan L, Yang G, Hong D (2020) Genetic dissection of thousand-seed weight and fine mapping of cqSW.A03-2 via linkage and association analysis in rapeseed (*Brassica napus* L.). Theor Appl Genet 133(4):1321–1335. https://doi.org/10.1007/s00122-020-03553-9
31. Wei D, Cui Y, Mei J, Qian L, Lu K, Wang ZM, Li J, Tang Q, Qian W (2019) Genome-wide identification of loci affecting seed glucosinolate contents in *Brassica napus* L. J Integr Plant Biol 61(5):611–623. https://doi/10.1111/jipb.12717
32. Wu D, Liang Z, Yan T, Xu Y, Xuan L, Tang J, Zhou G, Lohwasser U, Hua S, Wang H, Chen X, Wang Q, Zhu L, Maodzeka A, Hussain N, Li Z, Li X, Shamsi IH, Jilani G, Wu L, Zheng H, Zhang G, Chalhoub B, Shen L, Yu H, Jiang L (2019) Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. Mol Plant 12(1):30–43. https://doi.org/10.1016/j.molp.2018.11.007
33. Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide association studies in Maize: Praise and Stargaze. Mol Plant 10(3):359–374. https://doi.org/10.1016/j.molp.2016.12.008
34. Xiao Z, Zhang C, Tang F, Yang B, Zhang L, Liu J, Huo Q, Wang S, Li S, Wei L, Du H, Qu C, Lu K, Li J, Li N (2019) Identification of candidate genes controlling oil content by combination of genome-wide

association and transcriptome analysis in the oilseed crop *Brassica napus*. Biotechnol Biofuels 12:216. https://doi.org/10.1186/s13068-019-1557-x

35. Xiao Z, Tang F, Zhang L, Li S, Wang S, Huo Q, Yang B, Zhang C, Wang D, Li Q, Wei L, Guo T, Qu C, Lu K, Zhang Y, Guo L, Li J, Li N (2021) The *Brassica napus* fatty acid exporter FAX1-1 contributes to biological yield, seed oil content, and oil quality. Biotechnol Biofuels 14(1):190. https://doi.org/10.1186/s13068-021-02035-4
36. Xu R, Duan P, Yu H, Zhou Z, Zhang B, Wang R, Li J, Zhang G, Zhuang S, Lyu J, Li N, Chai T, Tian Z, Yao S, Li Y (2018) Control of Grain Size and Weight by the *OsMKKK10-OsMKK4-OsMAPK6* Signaling Path way in Rice. Mol Plant 11(6):860–873. https://doi.org/10.1016/j.molp.2018.04.004
37. Yang P, Shu C, Chen L, Xu J, Wu J, Liu K (2012) Identification of a major QTL for silique length and seed weight in oilseed rape (*Brassica napus* L.). Theor Appl Genet 125(2):285–296. https://doi.org/10.1007/s00122-012-1833-7
38. Yu F, Liang K, Fang T, Zhao H, Han X, Cai M, Qiu F (2019) A group VII ethylene response factor gene, *ZmEREB180*, coordinates waterlogging tolerance in maize seedlings. Plant Biotechnol J 17(12):2286–2298. https://doi.org/10.1111/pbi.13140
39. Zhang W, Xu W, Zhang H, Liu X, Cui X, Li S, Song L, Zhu Y, Chen X, Chen H (2021) Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. Theor Appl Genet 134(5):1329–1341. https://doi.org/10.1007/s00122-021-03774-6
40. Zhang X, Guan Z, Wang L, Fu J, Zhang Y, Li Z, Ma L, Liu P, Zhang Y, Liu M, Li P, Zou C, He Y, Lin H, Yuan G, Gao S, Pan G, Shen Y (2020) Combined GWAS and QTL analysis for dissecting the genetic architecture of kernel test weight in maize. Mol Genet Genomics 295(2):409–420. https://doi.org/10.1007/s00438-019-01631-2
41. Zhao B, Dai A, Wei H, Yang S, Wang B, Jiang N, Feng X (2016a) *Arabidopsis* KLU homologue *GmCYP78A72* regulates seed size in soybean. Plant Mol Biol 90(1–2):33–47. https://doi.org/10.1007/s11103-015-0392-0
42. Zhao W, Wang X, Wang H, Tian J, Li B, Chen L, Chao H, Long Y, Xiang J, Gan J, Liang W, Li M (2016b) Genome-wide identification of QTL for seed yield and yield-related traits and construction of a high-density consensus map for QTL comparison in *Brassica napus*. Front Plant Sci 7:17. https://doi.org/10.3389/fpls.2016.00017
43. Zheng M, Peng C, Liu H, Tang M, Yang H, Li X, Liu J, Sun X, Wang X, Xu J, Hua W, Wang H (2017) Genome-wide association study reveals candidate genes for control of plant height, branch initiation height and branch number in rapeseed (*Brassica napus* L.). Front Plant Sci 8:1246. https://doi.org/10.3389/fpls.2017.01246
44. Zhou Q, Han D, Mason AS, Zhou C, Zheng W, Li Y, Wu C, Fu D, Huang Y (2018) Earliness traits in rapeseed (*Brassica napus*): SNP loci and candidate genes identified by genome-wide association analysis. DNA Res 25(3):229–244. https://doi.org/10.1093/dnares/dsx052

Figures

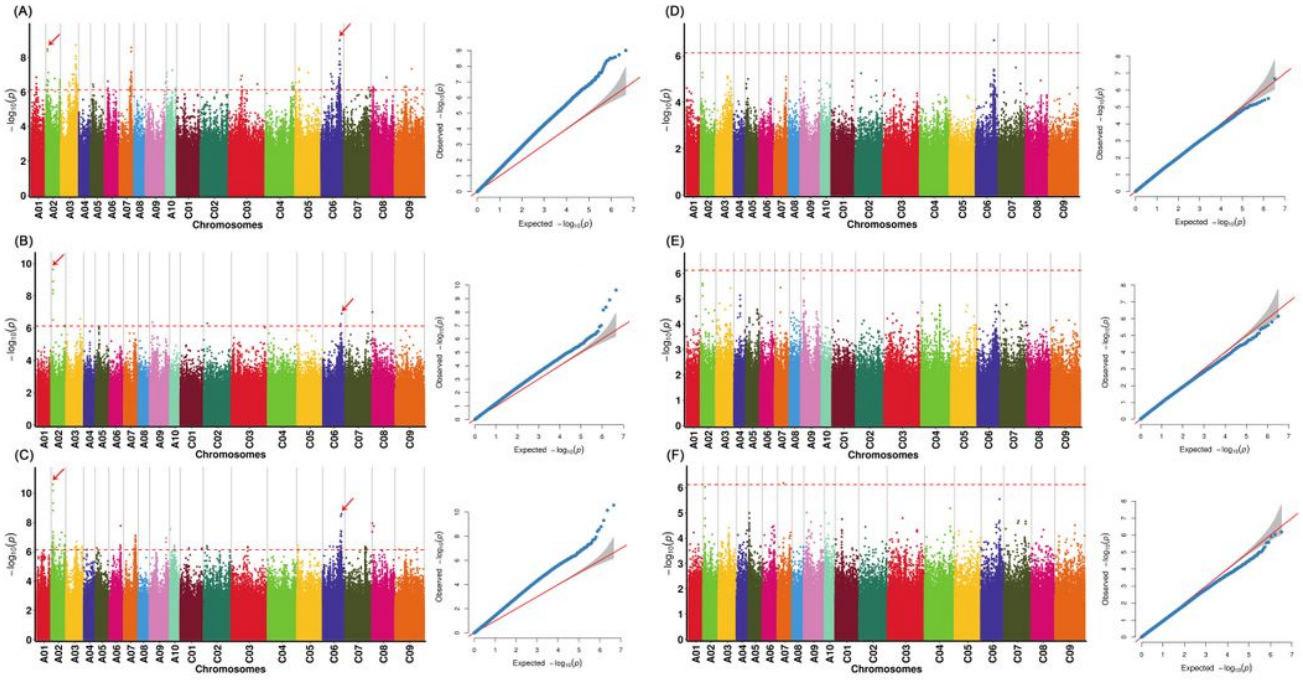

Figure 1

Genome wide association analysis for SW in an association panel of *B. napus*. (A) SW in Trial 1 under GLM model. (B) SW in Trial 2 under GLM model. (C) the mean value of SW in two trials under GLM model. (D) SW in Trial 1 under MLM model. (E) SW in Trial 2 under MLM model. (F) the mean value of SW in two trials under MLM model. SW, 1000-seed weight.

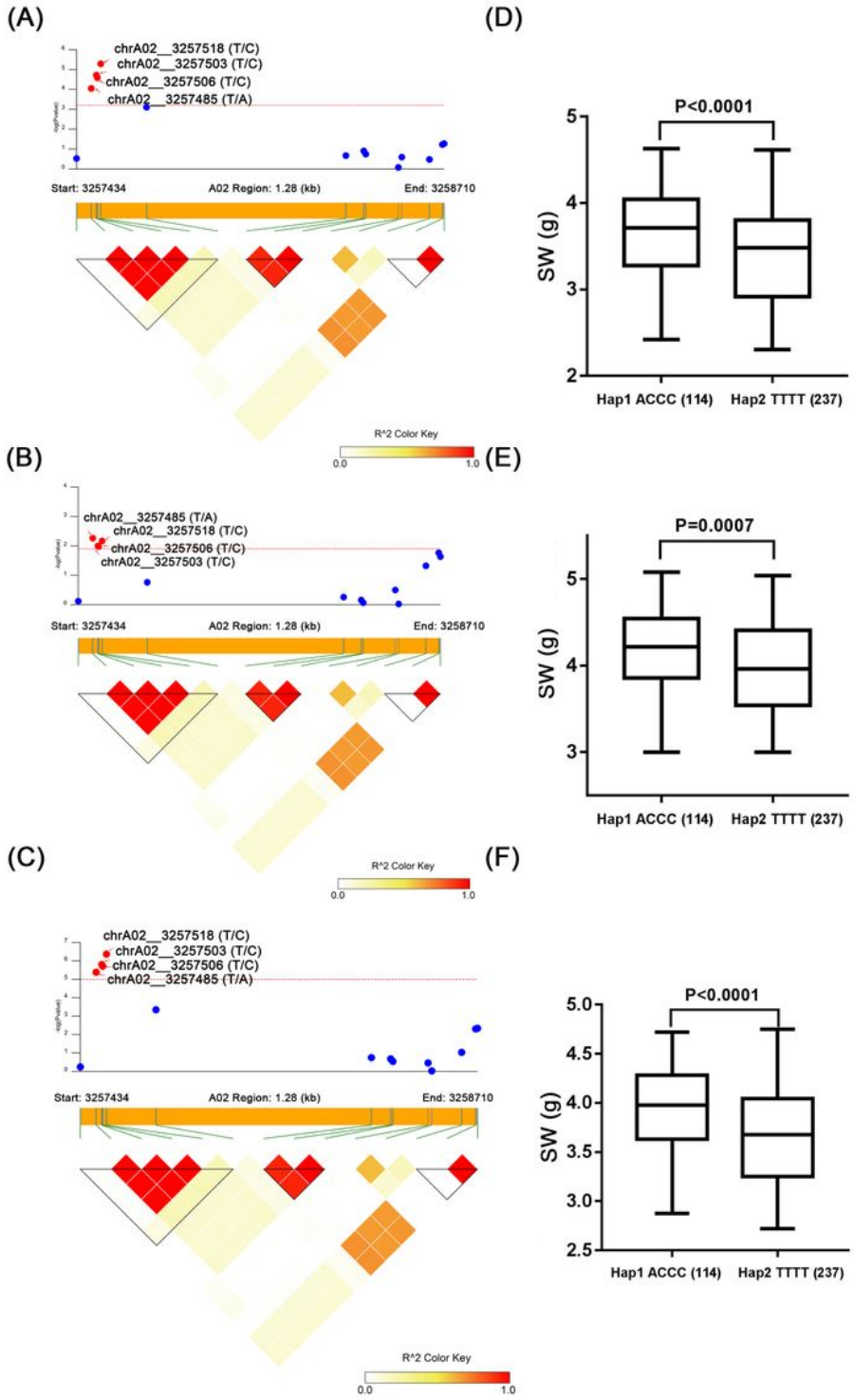

Figure 2

Candidate gene association analysis of *BnaA02g06870D* with SW. (A), Trial 1; (B) Trial 2 (C) the mean value of two trials. (D - F) Box plots for 1000-seed weight in two haplotypes of *BnaA02g06870D* in Trial 1 (D), Trial 2 (E) and the mean values of two trials (F). SW, 1000-seed weight. The number of inbred lines harbouring the corresponding allele is shown in brackets at the bottom.

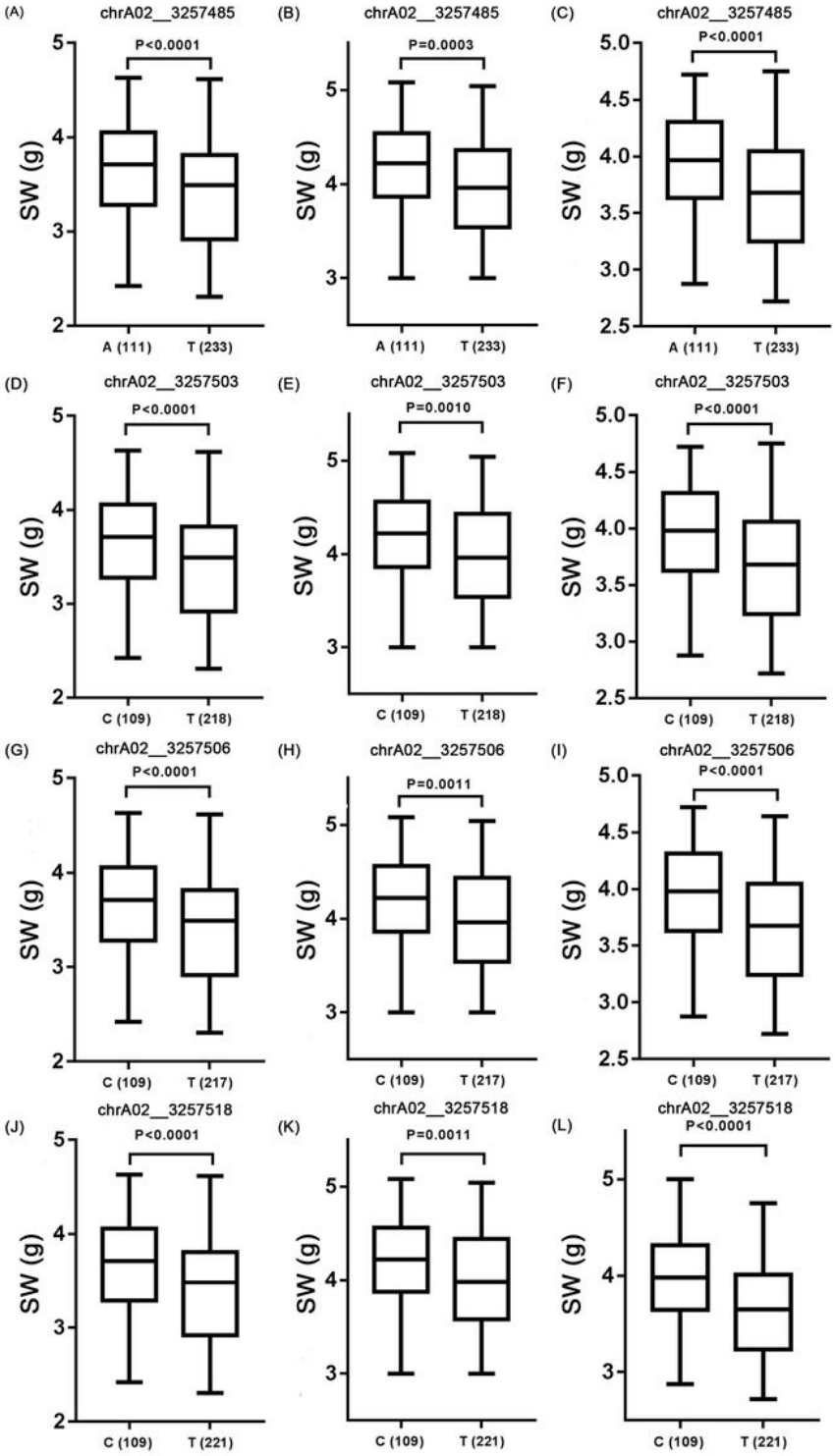

Figure 3

Association of alleles for SW. (A) chrA02__3257485 in Trial 1, (B) chrA02__3257485 in Trial 2, (C) chrA02__3257485 of the mean value in two trials, (D) chrA02__3257503 in Trial 1, (E) chrA02__3257503 in Trial 2, (F) chrA02__3257503 of the mean value in two trials, (G) chrA02__3257506 in Trial 1, (H) chrA02__3257506 in Trial 2, (I) chrA02__3257506 of the mean value in two trials, (J) chrA02__3257518 in Trial 1, (K) chrA02__3257518 in Trial 2 and (L) chrA02__3257518 of the mean value in two trials. SW,

1000-seed weight. The number of inbred lines harboring the corresponding allele are shown in the bracket at the bottom.

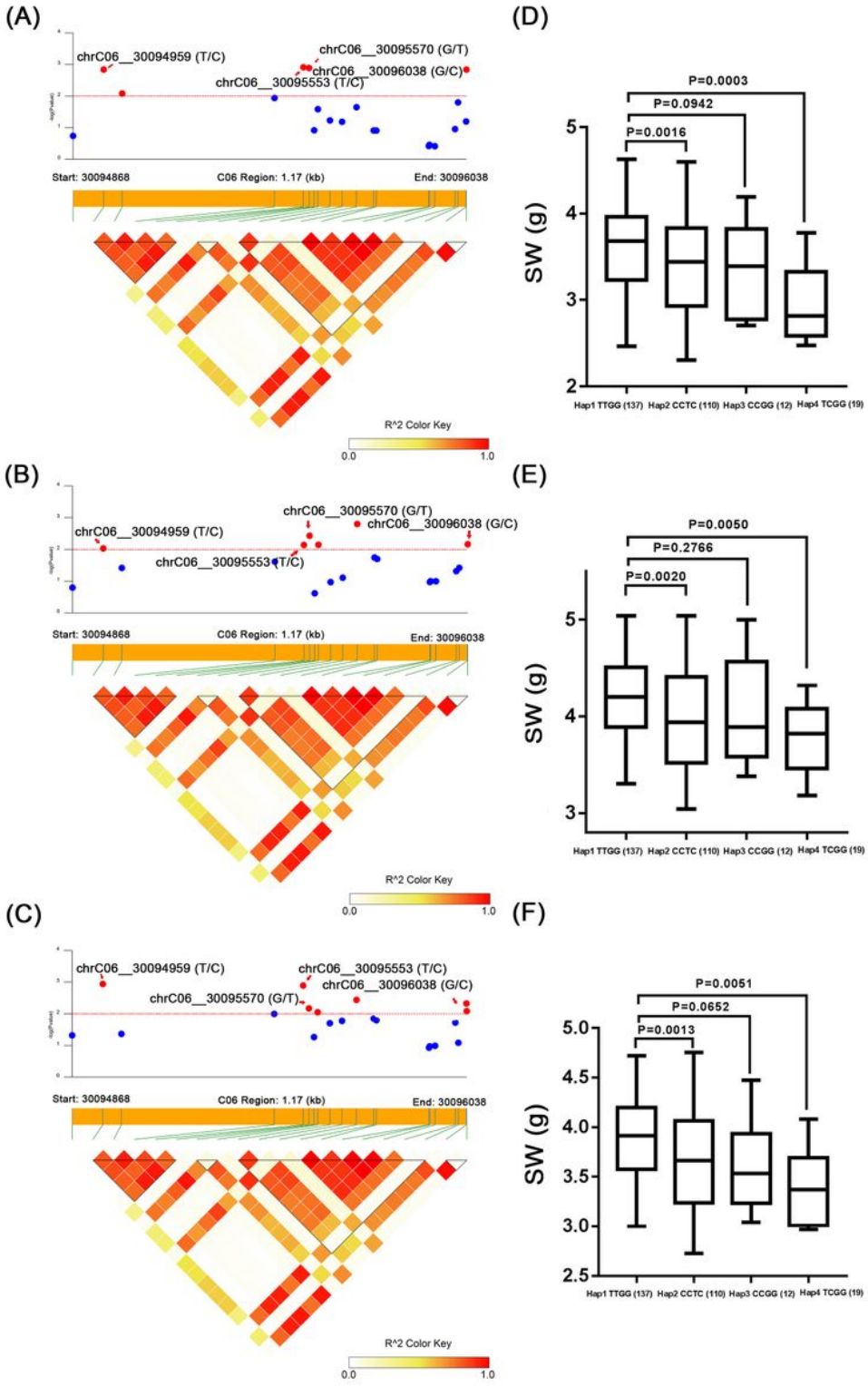

Figure 4

Candidate gene association analysis of *BnaC06g28920D* with SW. (A), Trial 1; (B) Trial 2; (C) the mean value of two trials. (D-F) Box plots for SW of four haplotypes of *BnaC06g28920D* in Trial 1 (D), Trial 2(E)

and the mean value of two trials (F). The number of inbred lines harbouring the corresponding allele is shown in brackets at the bottom. SW, 1000-seed weight.

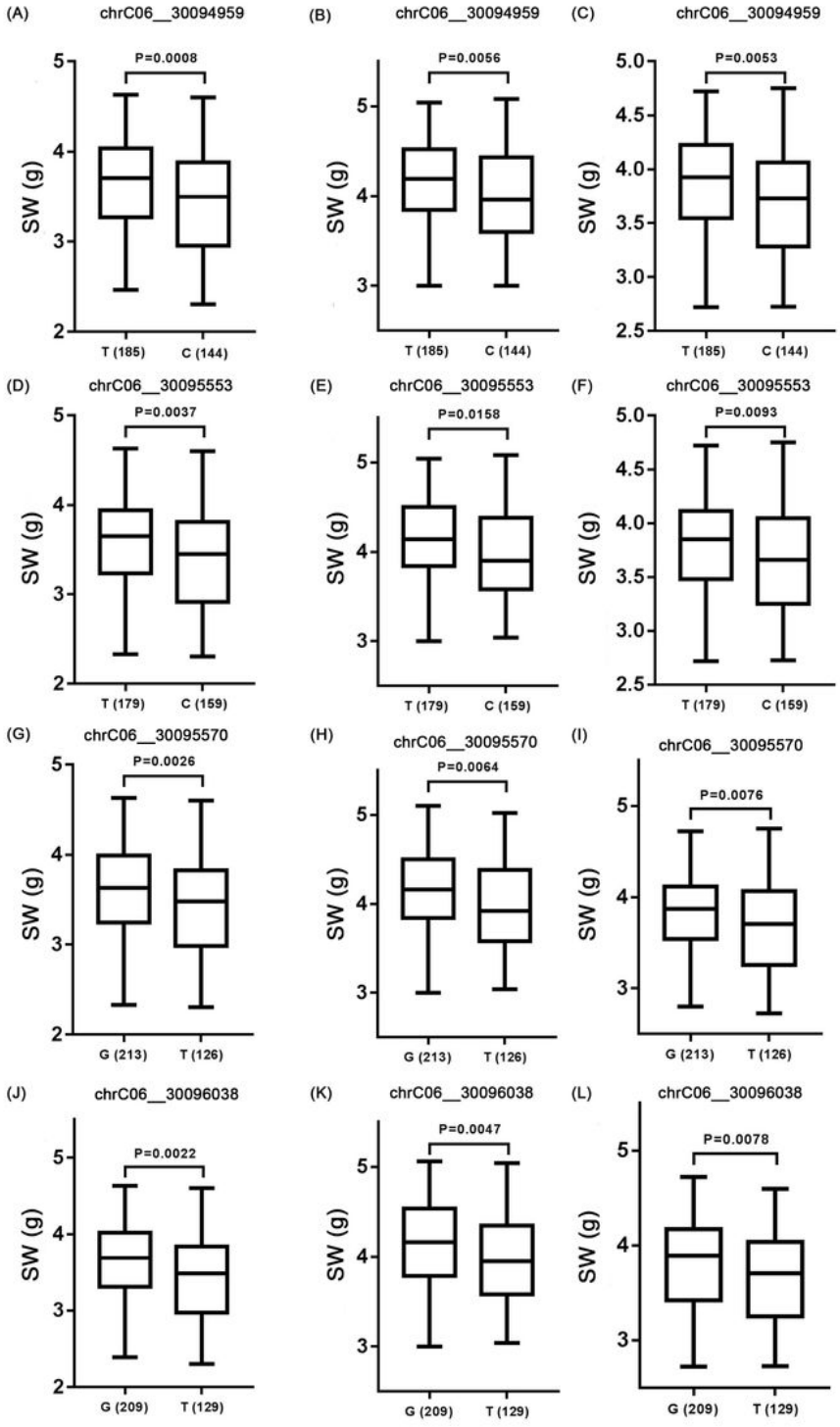

Figure 5

Association of alleles for SW. (A) chrC06__30094959 in Trial 1, (B) chrC06__30094959 in Trial 2, (C) chrC06__30094959 of the mean value in two trials, (D) chrC06__30095553 in Trial 1, (E)

chrC06__30095553 in Trial 2, (F) chrC06__30095553 of the mean value in two trials, (G) chrC06__30095570 in Trial 1, (H) chrC06__30095570 in Trial 2, (I) chrC06__30095570 of the mean value in two trials, (J) chrC06__30096038 in Trial 1, (K) chrC06__30096038 in Trial 2 and (L) chrC06__30096038 of the mean value in two trials. SW, 1000-seed weight. The number of inbred lines harboring the corresponding allele are shown in the bracket at the bottom.

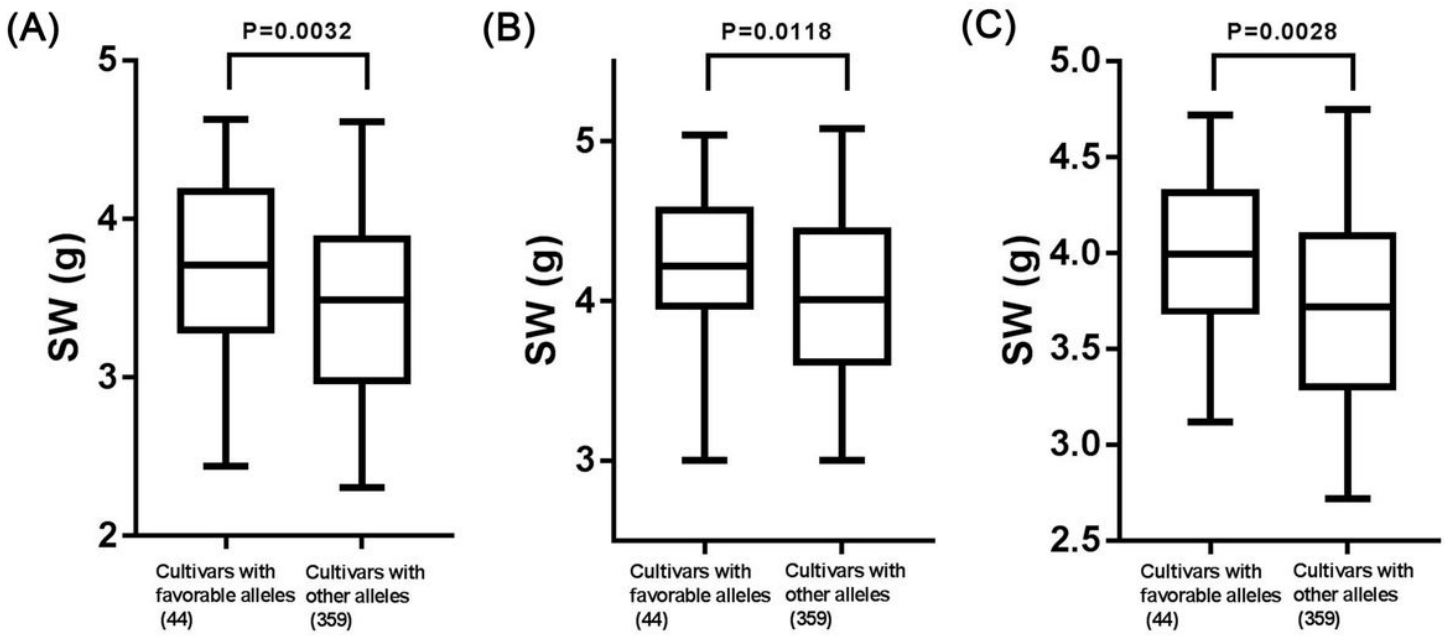

Figure 6

Differences in SW between *B. napus* cultivars with eight favorable alleles and *B. napus* cultivars with other alleles in the association panel. SW, 1000-seed weight.

Supplementary Files

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

- FigureS1.docx
- FigureS2.docx
- TableS1S6.xlsx
- TableS7.docx