Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide association study identifies candidate genes and favorable haplotypes for seed yield in Brassica napus

Haijiang Liu
Maoyan Zou
Bingbing Zhang
Xinyu Yang
Pan Yuan
Guangda Ding
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, seed yield, candidate genes association analysis, haplotype analysis

Posted Date: July 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1755224/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Oilseed rape (*Brassica napus* L.) is one of the most essential oil crops. Genetic improvement of seed yield (SY) is a major aim of *B. napus* breeding. Several studies have been reported the genetic mechanisms of SY of *B. napus.* Here, a genome-wide association study (GWAS) of SY was conducted using a panel of 403 natural accessions of *B. napus*, with more than five million high-quality single nucleotide polymorphisms (SNPs). A total of 1773 significant SNPs were detected associated with SY and 783 significant SNPs were co-located with previously reported QTLs. The lead SNPs chrA01__8920351 and chrA02__4555979 were jointly detected in Trial 2_2 and Trial 2_mean value, and in Trial 1_2 and Trial 1_mean value, respectively. Subsequently, two candidate genes of *BnaA01g17200D* and *BnaA02g08680D* were identified by combining transcriptome, candidate gene association analysis and haplotype analysis. *BnaA09g10430D* detected through lead SNP chrA09__5160639 was associated with SY of *B. napus*. Our results provide valuable information for studying the genetic control of seed yield in *B. napus* and can provide valuable genes, haplotypes and cultivars resources for the breeding of high seed yield *B. napus* cultivars.

Introduction

Oilseed rape (*Brassica napus* L.) is the second most important oil crop after soybean in the world (Zhang et al., 2022). Increasing seed yield (SY) of *B. napus* can meet the growing global demand for oil and protein (Zhao et al., 2016). Studying the genetic basis of SY in *B. napus* and mining candidate genes associated with SY has become particularly important (Raboanatahiry et al., 2018). In recent years, linkage mapping analysis has been widely applied to study the genetic basis of SY in *B. napus* (Shi et al., 2009; Zhao et al., 2016; Chao et al., 2019). Two hundred and twenty-six quantitative trait loci (QTLs) have been identified for yield and yield related traits (biomass yield, thousand-seed weight, plant height, first effective branch height, first effective branch number, length of main inflorescence, and pod number of main inflorescence) in eight environments (Zhao et al., 2016). Among them, eighteen QTLs were detected for SY, explaining 5.63%-20.91% of the phenotypic variation (PVEs) (Zhao et al., 2016). Besides that, a total of 870 QTLs were identified for nine yield related traits, of which nearly 10% (85) were identified for SY, which can explain 3.1–33.4% of the PVEs (Shi et al., 2009).

Compared with linkage analysis, GWAS has the advantages of high positioning accuracy without constructing mapping population (Xiao et al., 2017; Liu et al., 2022). GWAS has been widely applied to the genetic dissection of seed (grain) yield in different crops, such as rice (Zhong et al., 2021), maize (Ma and Cao, 2021), oil palm (Kalyana et al., 2021), wheat (Garcia et al., 2019), soybean (Hu et al., 2020) and cotton (Zhu et al., 2021). In rice, seventy-four quantitative trait nucleotides (QTN) were significant associated with grain yield and yield related traits, such as, grain length (GL), grain width (GW), grain thickness (GT), thousand-seed weight (TSW), and yield per plant (YYP), respectively, and 17 QTNs for YYP explained 0.28%-6.99% of the PVEs (Zhong et al., 2021). The association panel of maize consisting of 309 cultivars were used to perform GWAS for eight grain yield traits (Ma and Cao, 2021). Only one SNP was identified for grain yield per plant with a PVE of 5.92% (Ma and Cao, 2021). In recent years, GWAS

have also been widely used in the genetic analysis of complex agronomic traits in *B. napus*, including harvest index (Lu et al., 2017), plant height (Sun et al., 2016), branch number (Zheng et al., 2017), branch angle (Shen et al., 2018) and seed yield (Pal et al., 2021). A total of 16, 18, 27 and 18 SNPs were identified to be significantly associated for siliqua length, number of seeds per siliqua, thousand-seed weight and SY, respectively (Pal et al., 2021). In addition, six candidate genes for SY were identified on chromosomes A02 and C02 (Pal et al., 2021). However, the genotypes used in above GWAS were all based on 60 K SNP chips. With the continuous development of sequencing technology, genetic markers developed by whole genome re-sequencing have been widely used in *B. napus* GWAS analysis (Tang et al., 2021).

In this study, we investigated the SY in an association panel of *B. napus* across two years. GWAS of SY was performed using more than five million SNPs by whole-genome resequencing. A total of 1773 significant SNPs were identified to be associated with SY and three candidate genes were identified. In addition, three haplotypes of *BnaA01g17200Hap1* (TTAT), *BnaA02g08680Hap1* (GGCT) and *BnaA09g10430Hap2* (TGACATCCCCG) associated with high SY were revealed, which provided the molecular basis for subsequent molecular marker-assisted breeding.

Materials And Methods

Plant materials and field Trials

A total of 403 *B. napus* accessions were used in this study. Detailed information on them and their genotypes are described in our previous study (Liu et al., 2021). All accessions were planted in sandy loam soil at Meichuan Town, Wuxue city, Hubei province, China (115.55°E, 29.85°N) during *B. napus* growing seasons from 2018 to 2019 (Trial 1) and from 2019 to 2020 (Trial 2). Each accession had four rows and each plot had eight plants in each row. After harvest, six plants in each plot were selected to measure SY. The mean value, maximum, minimum and coefficient of variation were calculated using Excel 2007. Trial 1_mean value and Trial 2_mean value were calculated with the phenotypic values of three replications in Trial 1 and Trial 2, respectively (Supplementary Table 1). The R language was used to calculate the correlation coefficients between Trials.

Genome-wide association analysis and candidate gene identification

More than 5 million high-quality SNPs of the association panel were derived from previous study (Tang et al., 2021; Liu et al., 2021). Tassel 5.0 software was used for GWAS analysis by general linear models (GLM) and mixed linear models (MLM). The threshold for this population was 6.25×10^{-07} (Liu et al., 2021). GGplot2 (https://cran.r-project.org/web/packages/ggplot2/index.html) and CMplot (https://github.com/YinLiLin/CMplot) software were used to draw Manhattan plot and Quantile–Quantile plot, respectively. The population could be divided into five subgroups based on the cross validation errors and the r pairwise relative kinship was close to 0 (Liu et al., 2021). The LD value of the association panel is 300 kb (Liu et al., 2021), thus the genes located within LD decay (300 kb) value up/downstream of the peak SNPs were considered as candidate genes. The genotypes of *BnaA01g17200D*,

BnaA02g08680D and *BnaA09g10430D* in the association panel of *B. napus* were obtained by vcftools software (https://vcftools.github.io/index.html). Candidate gene association study of *BnaA01g17200D*, *BnaA02g08680D* and *BnaA09g10430D* 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 SY of the association panel of *B. napus*.

Candidate genes identification and haplotype analysis

The published genome of "Darmor-bzh" was used as a reference genome to annotate the candidate genes by BLAST analysis using gene sequence from *Arabidopsis* and rice. The finest hits for *B. napus* were considered to the homologous gene. To detected candidate genes for SY, all genes located in the 300 kb of up/downstream of candidate lead SNP were selected and common differentially expressed genes were predicted to be candidate genes combined with the transcriptome data of four high and four low yield *B. napus* cultivars (Lu et al., 2017). The annotation of candidate genes was obtained from TAIR10 (https://www.arabidopsis.org/). 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 SY among the haplotypes.

Results

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

SY ranged from 1.51 to 25.07 g/ plant (16.60-fold) in Trial 1 and from 2.18 to 35.88 g/ plant (16.45-fold) in Trial 2 (Table 1). A large range of variations were detected, with the coefficient of variation (CV) varying from 33.71% in Trial 2 to 46.12% in Trial 1 (Table 1). SY in both trials showed an approximate normal distribution (Supplementary Fig. 1). The correlation coefficient (r) of SY between Trial 1 and Trial 2 were 0.41 (Table 1, Supplementary Fig. 2).

Table 1
Mean, maximum (max), minimum (min), range and coefficient of variation (CV, %) of the seed yield (SY) in an association panel of *B. napus* in Trial 1 and Trial 2

	Seed yield (g/plant)				CV (%)
Trials	**Mean**	**Max**	**Min**	**Range**	
Trial 1_1	5.68	13.38	1.51	11.87	48.75%
Trial 1_2	9.92	21.71	1.85	19.86	44.19%
Trial 1_3	12.31	25.07	3.1	21.97	38.63%
Trial 1_mean value	9.34	23.35	1.53	21.83	32.49%
Trial 2_1	17.06	34.47	2.18	32.39	42.85%
Trial 2_2	16.31	31.15	2.85	28.30	36.65%
Trial 2_3	16.96	35.88	2.18	33.7	44.09%
Trial 2_mean value	16.79	35.79	2.4	33.38	32.86%

GWAS of SY of B. napus

GLM and MLM approaches were used to perform GWAS analysis to reveal SNPs significantly associated with SY of *B. napus* (Supplementary Fig. 3–4, Supplementary Table 2–3). A total of 1773 SNPs were identified to be significantly associated with SY across two trials by GLM approach ($P < 6.25 \times 10^{-07}$) (Supplementary Table 2). The PVE of 1773 SNPs ranged from 5.57–12.73% (Supplementary Table 2). Thirteen, 20, 19, 34, 38, 92, 390 and 1167 significant SNPs were associated with SY in Trial 1_1, Trial 1_2, Trial 1_3, Trial 1_mean value, Trial 2_1, Trial 2_2, Trial 2_3 and Trial 2_mean value, respectively, and distributed on 19 *B. napus* chromosomes (Supplementary Fig. 3, Supplementary Table 2). Chromosomes C09 and A09 had the largest (404) and the second largest (281) number of significant SNPs, respectively (Supplementary Table 2). Both MLM and GLM analysis identified the same 16 significant SNPs located on chromosomes A02, A05, A09, A10, C03, C04, C06, explaining 9.78–12.08% of PVE (Supplementary Table 3).

Candidate gene association analysis and haplotype analysis of BnaA01g17200D **and** BnaA02g08680D

The SNP of chrA01__8920351 co-located on chromosome A01 were significantly associated with SY in the both Trial 2_2 and Trial 2_mean value (Fig. 1a, Supplementary Table 2, 4). Within the 300 kb up/downstream of chrA01__8920351, a total of 87 genes were found (Supplementary Table 5). Previously, transcriptome sequencing was performed on extremely high and low-yield *B. napus* cultivars to mine yield related candidate genes (Lu et al., 2017). We compared the GWAS results in this study with previous transcriptome data and results show that *BnaA01g17200D* was the only one significantly differentially expressed gene among the 87 candidate genes on A01 chromosome (Table 2,

Supplementary Table 5). The expression level of *BnaA01g17200D* in bud on the primary branch of high-yield cultivars was significantly higher than that of low-yield cultivars (Table 2, Supplementary Table 5). We extracted the SNPs within the coding region and promoters (upstream, 2000 bp) of *BnaA01g17200D*, and eight SNPs were located within the 2 kb promoter region and the entire coding region of *BnaA01g17200D* (Supplementary Table 6). Four SNPs in *BnaA01g17200D* were detected to be significantly associated with SY ($P < 0.001$) (Table 3). Further analyzed the four significant SNPs showed that, the T allele of chrA01__9042710, T allele of chrA01__9042762, A allele of chrA01__9044014 and T allele of chrA01__9044072 were high-seed yield alleles (Fig. 1b-f). According to the above four SNPs, two major haplotypes were detected, and cultivars with *BnaA01g17200Hap1* (TTAT) had much higher SY than cultivars with *BnaA01g17200Hap2* (CAGG) (Fig. 1g).

Table 2
Common candidate genes identified by GWAS and transcriptome sequencing

Gene ID	Log_2 (fold change)	P value	*Arabidopsis* homologue	Annotations	Expression sites
BnaA01g17200D	1.97	5.62E-16	*AT4G27140*	seed storage albumin 1	Bud on the primary branch
BnaA02g08680D	-6.14	1.05E-08	*AT2G30480*	hypothetical protein	Bud on the main inflorescence

Table 3
List of synonymous and non-synonymous SNP variants identified in the candidate genes of *BnaA01g17200D*, *BnaA02g08680D* and *BnaA09g10430D*.

Candidate genes	SNP	Major allele	Minor allele	SNP types	Amino acid changes
BnaA01g17200D	chrA01__9042710	T	C	Promoter	\
	chrA01__9042762	T	A	Promoter	\
	chrA01__9044014	A	G	Promoter	\
	chrA01__9044072	T	C	Promoter	\
BnaA02g08680D	chrA02__4290037	G	A	Promoter	\
	chrA02__4290087	G	A	Promoter	\
	chrA02__4290091	C	T	Promoter	\
	chrA02__4290092	T	A	Promoter	\
BnaA09g10430D	chrA09__5322341	C	T	Exon	Arginine to Cysteine
	chrA09__5322518	C	T	Intron	\
	chrA09__5322553	T	C	Intron	\
	chrA09__5322559	C	G	Intron	\
	chrA09__5322560	T	G	Intron	\
	chrA09__5322565	A	G	Intron	\
	chrA09__5322568	G	A	Intron	\
	chrA09__5322918	C	T	Intron	\
	chrA09__5322966	A	G	Intron	\
	chrA09__5323050	A	G	Exon	Isoleucine to Valine
	chrA09__5323064	T	A	Exon	Leucine to Terminator
	chrA09__5323081	A	C	Exon	Glutamine to Histidine
	chrA09__5323251	T	A	Promoter	\
	chrA09__5323456	C	T	Promoter	\
	chrA09__5323496	T	C	Promoter	\
	chrA09__5323633	G	C	Promoter	\
	chrA09__5323759	A	C	Promoter	\

Candidate genes	SNP	Major allele	Minor allele	SNP types	Amino acid changes
	chrA09__5324271	T	C	Promoter	\
	chrA09__5324376	A	G	Promoter	\

The SNP of chrA02__4555979 on chromosome A02 was identified in Trial 1_2 and Trial 1_mean value (Fig. 2a, Supplementary Table 2, 4). Based on the LD decay (300 kb), 108 genes were selected within the 300 kb up/downstream of the chrA02__4555979 (Supplementary Table 6). The expression of *BnaA02g08680D* was 6.14-fold between high-yield *B. napus* cultivars and low-yield *B. napus* cultivars in bud on the main inflorescence (Lu et al., 2017, Table 2). Ten SNPs were extracted within the coding region and promoters (upstream, 2000 bp) of *BnaA02g08680D* and four SNPs in *BnaA02g08680D* were significantly associated with the SY, and the G allele of chrA02__4290037, G allele of chrA02__4290087, C allele of chrA02__4290091 and T allele of chrA02__4290092 were the high-seed yield alleles (Fig. 2b-f, Supplementary Fig. 6). The above significant SNPs revealed two major haplotypes, and cultivars harboring *BnaA02g08680Hap1* (GGCT) had significantly greater seed yield than cultivars harboring *BnaA02g08680Hap2* (AATA) (Fig. 2g).

Comparison Of The Significant Snps Associated With Sy With Qtls For Sy

Based on the Darmor-bzh reference genome, a total of 783 SNPs were co-located with the intervals of 47 QTLs for SY in the previous linkage analysis of the *B. napus* segregation populations (Bouchet et al., 2014; Cai et al., 2016; Luo et al., 2017; Supplementary Table 8). A significant SNP (chrA09__5160639) identified in Trial 2_3 was co-located with the interval of '*qSY38'* QTL and a gene (*BnaA09g10430D*) reported to control SY was within the interval of this QTL (Luo et al., 2017, Supplementary Table 8). *OsNPF7.2*, the orthologue of *BnaA09g10430D*, regulates tiller number and grain yield in rice (Wang et al., 2018). Candidate gene association analysis of *BnaA09g10430D* with SNP from the 2-kb promoter and the entire coding region showed that 19 SNPs were significantly associated with SY, and had strong LD with each other (Fig. 3b, Table 3, Supplementary Table 6). Among them, 7 were located in the promoter, 4 in the exon and 8 in the intron region (Table 3). Notably, chrA09__5322341 (C/T), chrA09__5323050 (A/G), chrA09__5323064 (T/A) and chrA09__5323081 (A/C) were located in the exon region of the gene *BnaA09g10430D* and resulted in amino acid changes from arginine to cysteine, isoleucine to valine, leucine to terminator and glutamine to histidine, respectively (Table 3). Further analysis demonstrated that the T allele of chrA09__5322341 and G allele of chrA09__5323050, A allele of chrA09__5323064, C allele of chrA09__5323081, A allele of chrA09__5323251, T allele of chrA09__5323456, C allele of chrA09__5323496, C allele of chrA09__5323633, C allele of chrA09_5323759, T allele of chrA09__5324271, and G allele of chrA09_5324376 were associated with high SY (Fig. 3c-m). Two major haplotype groups were obtained according to the above 11 significant SNPs (Fig. 3n). Among them, *BnaA09g10430Hap2* (TGACATCCCCG) had significantly higher SY than *BnaA09g10430Hap1* (CATATCTGATA), therefore *BnaA09g10430Hap2* was considered as the favorable haplotype (Fig. 3n). In

addition, a total of 67 *B. napus* cultivars contained these three favorable haplotypes had higher SY than other *B. napus* cultivars (Fig. 4).

Discussion

Overlapped loci for SY between this study and previous studies

In this study, SY of the panel of 403 *B. napus* accessions varied from 1.51 to 25.07 g/plant in Trial 1 and from 2.18 to 35.88 g/ plant in Trial 2 (Table 1, Supplementary Table 1). Compared with the association panel with 200 *B. napus* accessions by Pal et al. (2020), the natural population (403 *B. napus* accessions) panel in this study has more extensive phenotypic variation. In addition, more than 5 million SNPs by whole genome-wide re-sequencing were used for GWAS analysis in this study (Tang et al., 2021; Liu et al., 2021). Compared with previous studies using 60K SNP chip as genotype for GWAS of SY of *B. napus*, more significant SNPs related to SY were identified in this study (Supplementary Table 2–3).

Understanding the genetic variation of SY and SY related traits is helpful for improving crop yield. GWAS of SY related traits including PH (Sun et al., 2016), BN (Zheng et al., 2017; Liu et al., 2021), PB (Li et al., 2020), biological yield (Xiao et al., 2021) and SY in *B. napus* (Pal et al., 2021) have been conducted in recent years. For instance, a panel of 476 inbred lines of *B. napus* were genotyped using 19167 high-quality SNPs, and a total of 48 significant SNPs were identified associated with plant height (Sun et al., 2016). Using a panel of 331 *B. napus* accessions and 24508 SNPs, 27 loci were significantly associated with silique number (Li et al., 2020). In another study, a panel of 505 inbred lines of *B. napus* were genotyped using more than ten million SNPs, and a total of 88 significant loci were identified significant associated with biological yield (Xiao et al., 2021). Chloroplast membrane protein-located FATTY ACID EXPORTER 1–1 (FAX1-1) is identified as a candidate gene for the significant SNPs associated with biological yield by the combination of GWAS with transcriptome sequencing (Xiao et al., 2021). In this study, a total of 103 SNPs associated with SY were co-located with previously reported SNPs and QTLs for seed number per pod (49), pod number (21) and TSW (32) (Shi et al., 2015; Lu et al., 2017; Khan et al., 2019; Pal et al., 2021; Supplementary Table 9). In addition, two SNPs (C03.27718223 and C02.22109942) associated with SY identified by Pal (2021) were also identified in our study (Supplementary Table 8). These repeatedly identified significant SNPs should be focused in the further study.

In this study, a total of 1773 SNPs were detected for SY and 783 SNPs were co-located with QTLs for SY and SY-related QTLs by linkage analysis reported in previous studies (Bouchet et al., 2014; Luo et al., 2017; Cai et al., 2016; Pal et al., 2021) (Supplementary Table 8). ChrA09__5160639 was co-located with QTL "*qSY38*" (Luo et al., 2017) and *BnaA09g10430D* (*BnaA09.NPF8.5*) was considered to be the most likely candidate gene. *BnaA09.NPF8.5* had an MFS (Major facilitator superfamily) domain-contain protein. In rice, *NPF7.2* enhanced influx of nitrate and nitrate concentration and improved root length, root weight and fresh weight, thereby increasing tiller number and grain yield (Wang et al., 2018).

The candidate genes for SY in *B. napus*

In this study, 9.53% (169/1773) of the significant SNPs were identified in more than one repetition (Supplementary Table 4). For example, chrA02__4555979 was identified simultaneously in Trial 1_2 and Trial 1_mean value. *BnaA02g08680D* was located 269 kb upstream of chrA02__4438592. The expression of *BnaA02g08680D* in bud on the main inflorescence of high-yield cultivars was significantly higher than that of low-yield cultivars (Table 2). In addition, chrA01__8920351 was detected simultaneously in Trial 2_2 and Trial 2_mean value (Supplementary Table 2). *BnaA01g17200D* was located 121 kb downstream of chrA01__893123. The expression level of *BnaA01g17200D* in bud on the primary branch of high-yield cultivars was 6 times higher than that in low-yield cultivars (Table 2). In *Arabidopsis*, *ATSESA1* (SEED STORAGE ALBUMIN 1), the orthologue of *BnaA01g17200D*, regulates seed protein accumulation to influence seed growth (Hu et al., 2021). In addition, some genes which have been reported to improve SY were also identified in our GWAS (Supplementary Table 10). The mediator subunit, *OsMED15* is important for flowering, seed size and weight in rice (Dwivedi et al., 2019). The significant SNP chrA05__18407128 on A05 was identified to be located downstream 101 kb of *B. napus BnaA05.TCP4* (*TCP family transcription factor 4*), which is the homologous of *OsMED15*. Optimal flowering time is associated with high SY, and *AtTCP4* regulates flowering time in *Arabidopsis* (Schiessl et al., 2015; Kubota et al., 2017). *BnaC08g36050D*, a member of the AUX1 LAX family of auxin influx carriers, was located within the interval of SNP chrC08__33833529. *AT2G21050*, the homologous of *BnaC08g36050D* in *Arabidopsis*, has been reported to increase xylem area and enhance SY (Cabello et al., 2019). These reported SY related genes detected in this study further prove the accuracy of our GWAS results.

Favorable alleles, haplotypes and high-seed yield breeding in *B. napus*

Traditional breeding methods are difficult to meet the demand for food and edible oil caused by the growing population (Ma and Cao, 2021). Mining superior alleles and haplotypes for molecular marker assisted breeding has gradually become a new method for breeders (Ma and Cao, 2021). Candidate gene association study can detect genetic variants within genes that are significantly associated with phenotypes. Genetic variations are located in the promoter and coding regions, which may lead to the difference of gene expression level and protein sequences among cultivars, respectively. In this study, we conducted candidate gene association study on three candidate genes (*BnaA01g17200D*, *BnaA02g08680D* and *BnaA09g10430D*) and identified several genetic variations significantly related to SY within the genes. Four significant SNPs (chrA01__9042710, chrA01__9042762, chrA01__9044014 and chrA01__9044072) in *BnaA01g17200D* and four SNPs (chrA02__4290037, chrA02__4290087, chrA02__4290091 and chrA02__4290092) in *BnaA02g08680D* were all located in the promoter. Seven, four and eight significant SNPs associated with SY were located in the promoter, the exon and the intron of *BnaA09g10430D*, respectively; and the SNPs located in the exon caused the changes of amino acids among different haplotypes (Table 3). Three high-yield haplotypes (*BnaA01g17200Hap1*, *BnaA02g08680Hap1* and *BnaA09g10430Hap2*) were identified based on the above significant SNPs (Fig. 1–3). SY of 67 *B. napus* cultivars harboring these three favorable haplotypes were significantly

higher than that of other cultivars in two-year field experiments (Fig. 4). These favorable alleles and haplotypes will be valuable for the cultivation of high-yield *B. napus* cultivars.

Conclusion

SY of *B. napus* is a complex quantitative trait controlled by multiple genes. A total of 1773 SNPs were identified to be significantly associated with seed yield across two years, and 783 of them were co-located with previous reported QTLs. Candidate genes association and haplotype analysis of *BnaA01g17200D*, *BnaA02g08680D* and *BnaA09g10430D* revealed several high-yield haplotypes. These significant SNP loci, candidate genes, favorable alleles and haplotypes will be useful for the breeding of high-yield *B. napus* cultivars

Declarations

Author contribution HL, MZ and LS designed research, reviewed writing and drafted the manuscript. HL, MZ, BZ, XY and PY: participated the experiments. LS, GD and FX: Participate in manuscript revision

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

Data availability The data sets supporting the results of this article are included within the article and its additional files.

Code availability Not applicable

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of Interest Statement The authors declare no conflict of interest.

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 SRP072900 (Lu et al. 2017).

References

1. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. http://doi:10.1093/bioinformatics/bth457
2. Bouchet AS, Nesi N, Bissuel C, Bregeon M, Lariepe A, Navier H, Ribière N, Orsel M, Grezes-Besset B, Renard M, Laperche A (2014) Genetic control of yield and yield components in winter oilseed rape (*Brassica napus L.*) grown under nitrogen limitation. Euphytica 199:183–205. doi:10.1007/s10681-014-1130-4
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 23:2633–2635. http://doi:10.1093/bioinformatics/btm308
4. Cabello JV, Chan RL (2019) Arabidopsis and sunflower plants with increased xylem area show enhanced seed yield. Plant J 99:717–732. http://doi:10.1111/tpj.14356
5. Cai GQ, Yang QY, Chen H, Yang Q, Zhang CY, Fa CC, Zhou YM (2016) Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. Sci Rep 6:21625. http://doi:10.1038/srep21625
6. Chao HB, Raboanatahiry N, Wang XD, Zhao WG, Chen L, Guo LX, Li BJ, Hou DL, Pu S, Zhang LN, Wang H, Wang BS, Li MT (2019) Genetic dissection of harvest index and related traits through genome-wide quantitative trait locus mapping in *Brassica napus L*. Breed Sci 69:104–116. http://doi:10.1270/jsbbs.18115
7. 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 1862:194432. http://doi:10.1016/j.bbagrm.2019.194432
8. Garcia M, Eckermann P, Haefele S, Satija S, Sznajder B, Timmins A, Baumann U, Wolters P, Mather DE, Fleury D (2019) Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum L.*) evaluated in southern Australia. PLoS ONE 14:e0211730. http://doi:10.1371/journal.pone.0211730
9. Hu DZ, Zhang HR, Du Q, Hu ZB, Yang ZY, Li X, Wang J, Huang F, Yu DY, Wang H, Kan GZ (2020) Genetic dissection of yield-related traits via genome-wide association analysis across multiple environments in wild soybean (*Glycine soja Sieb. and Zucc.*). Planta 251:39. http://doi:10.1007/s00425-019-03329-6
10. Hu YL, Zhou LM, Yang YH, Zhang WB, Chen ZH, Li XM, Qian Q, Kong FJ, Li YG, Liu X, Hou XL (2021) The gibberellin signaling negative regulator RGA-LIKE3 promotes seed storage protein accumulation. Plant Physiol 185:1697–1707. http://doi:10.1093/plphys/kiaa114
11. Kalyana BB, Mathur RK, Venu MVB, Sandip S, Ravichandran G, Anita P, Bhagya HP (2021) Genome-wide association study (GWAS) of major QTLs for bunch and oil yield related traits in *Elaeis guineensis L*. Plant Sci 305:110810. http://doi:10.1016/j.plantsci.2020.110810
12. Kubota A, Ito S, Shim JS, Johnson RS, Song YH, Breton G, Goralogia GS, Kwon MS, Cintrón DL, Koyama T, Ohme-Takagi M, Pruneda-Paz JL, Kay SA, MacCoss MJ, Imaizumi T (2017) TCP4-

dependent induction of CONSTANS transcription requires GIGANTEA in photoperiodic flowering in *Arabidopsis*. PLoS Genet 13:e1006856. http://doi:10.1371/journal.pgen.1006856

13. Khan SU, Jiao YM, Liu S, Zhang KP, Khan MHU, Zhai YG, Olalekan A, Fan CC, Zhou YM (2019) Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in *Brassica napus L*. Ind Crops Prod 142:111877. https://doi.org/10.1016/j.indcrop.2019.111877
14. Li SY, Zhu YY, Varshney RK, Zhan JP, Zheng XX, Shi JQ, Wang XF, Liu GH, Wang HZ (2020) A systematic dissection of the mechanisms underlying the natural variation of silique number in rapeseed (*Brassica napus L*.) germplasm. Plant Biotechnol J 18:568–580. http://doi:10.1111/pbi.13224
15. Liu HJ, Wang JC, Zhang BB, Yang XY, Hammond JP, Ding GD, Wang SL, Cai HM, Wang C, Xu FS, Shi L (2021) Genome-wide association study dissects the genetic control of plant height and branch number in response to low-phosphorus stress in *Brassica napus*. Ann Botany 128:919–930. http://doi:10.1093/aob/mcab115
16. Liu HJ, Wang W, Yang M, Yuan P, Hammond JP, King GJ, Ding GD, White PJ, Wang SL, Cai HM, Wang C, Lu CG, Xu FS, Shi L (2022) Genome-wide association studies of important agronomic traits in *Brassica napus*: What we have learned and where we are headed. Annual Plant Reviews 5:1–30. http://doi:10.1002/9781119312994.apr0801
17. Lu K, Peng L, Zhang C, Lu JH, Yang B, Xiao ZC, Liang Y, Xu XF, Qu CM, Zhang K, Liu LZ, Zhu QL, Fu ML, Yuan XY, Li JN (2017) Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in *Brassica napus*. Front Plant Sci 8:206. http://doi:10.3389/fpls.2017.00206
18. Luo ZL, Wang M, Long Y, Huang YJ, Shi L, Zhang CY, Liu X, Fitt BDL, Xiang JX, Mason AS, Snowdon RJ, Liu PF, Meng JL, Zou J (2017) Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. Theor Appl Genet 130:1569–1585. http://doi:10.1007/s00122-017-2911-7
19. Ma J, Cao YY (2021) Genetic dissection of grain yield of maize and yield-related traits through association mapping and genomic prediction. Front Plant Sci 12:690059. http://doi:10.3389/fpls.2021.690059
20. Pal L, Sandhu SK, Bhatia D, Sethi S (2021) Genome-wide association study for candidate genes controlling seed yield and its components in rapeseed (*Brassica napus subsp. napus*). Physiol Mol Biology Plants 27:1933–1951. http:// doi:10.1007/s12298-021-01060-9
21. Raboanatahiry N, Chao HB, Dalin H, Pu S, Yan W, Yu LJ, Wang BS, Li MT (2018) QTL alignment for seed yield and yield related traits in *Brassica napus*. Front Plant Sci 9:1127. http://doi:10.3389/fpls.2018.01127
22. Schiessl S, Iniguez-Luy F, Qian W, Snowdon RJ (2015) Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*. BMC Genomics 16:737. http://doi:10.1186/s12864-015-1950-1

23. Shen YS, Yang Y, Xu ES, Ge XH, Xiang Y, Li ZY (2018) Novel and major QTL for branch angle detected by using DH population from an exotic introgression in rapeseed (*Brassica napus L.*). Theor Appl Genet 131:67–78. http://doi:10.1007/s00122-017-2986-1
24. Shi JQ, Li RY, Qiu D, Jiang CC, Long Y, Morgan CL, Bancroft I, Zhao JY, Meng JL (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. Genetics 182:851–861. http://doi:10.1534/genetics.109.101642
25. Shi JQ, Zhan JP, Yang YH, Ye J, Huang SM, Li RY, Wang XF, Liu GH, Wang HZ (2015) Linkage and regional association analysis reveal two new tightly-linked major-QTLs for pod number and seed number per pod in rapeseed (*Brassica napus L.*). Sci Rep 5:14481. http://doi:10.1038/srep14481
26. Sun CM, Wang BQ, Yan L, Hu KN, Liu S, Zhou YM, Guan CY, Zhang ZQ, Li JN, Zhang JF, Chen S, Wen J, Ma CZ, Tu JX, Shen JX, Fu TD, Yi B (2016) Genome-wide association study provides insight into the genetic control of plant height in rapeseed (*Brassica napus L.*). Front Plant Sci 7:1102. http://doi:10.3389/fpls.2016.01102
27. Tang S, Zhao H, Lu SP, Yu LQ, Zhang GF, Zhang YT, Yang QY, Zhou YM, Wang XM, Ma W, Xie WB, Guo L (2021) Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. Mol Plant 14:470–487. http://doi:10.1016/j.molp.2020.12.003
28. Wang J, Lu K, Nie HP, Zeng QS, Wu BWJ, Qian J, Fang ZM (2018) Rice nitrate transporter OsNPF7.2 positively regulates tiller number and grain yield. Rice 11:12. http://doi:10.1186/s12284-018-0205-6
29. Xiao YJ, Liu HJ, Wu LJ, Warburton M, Yan JB (2017) Genome-wide association studies in maize: praise and stargaze. Mol Plant 10:359–374. http://doi:10.1016/j.molp.2016.12.008
30. Xiao ZC, Tang F, Zhang LY, Li ST, Wang SF, Huo Q, Yang B, Zhang C, Wang DJ, Li Q, Wei LJ, Guo T, Qu CM, Lu K, Zhang YF, Guo L, Li JN, Li NN (2021) The *Brassica napus* fatty acid exporter FAX1-1 contributes to biological yield, seed oil content, and oil quality. Biotechnol Biofuels 14:190. http://doi:10.1186/s13068-021-02035-4
31. Zhang ST, Lu JW, Zhu Y, Fang YT, Cong RH, Li XK, Ren T (2022) Rapeseed as a previous crop reduces rice N fertilizer input by improving soil fertility. Field Crops Research 281:108487. https://doi.org/10.1016/j.fcr.2022.108487
32. Zhao WG, Wang XD, Wang H, Tian JH, Li BJ, Chen L, Chao HB, Long Y, Xiang J, Gan JP, Liang WS, Li MT (2016) 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. http://doi:10.3389/fpls.2016.00017
33. Zheng M, Peng C, Liu HF, Tang M, Yang HL, Li XK, Liu JL, Sun XC, Wang XF, Xu JF, Hua W, Wang HZ (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. http://doi:10.3389/fpls.2017.01246
34. Zhong H, Liu S, Sun T, Kong WL, Deng XD, Peng ZH, Li YS (2021) Multi-locus genome-wide association studies for five yield-related traits in rice. BMC Plant Biol 21:364.

http://doi:10.1186/s12870-021-03146-8

35. Zhu GZ, Hou S, Song XH, Wang X, Wang W, Chen QJ, Guo WZ (2021) Genome-wide association analysis reveals quantitative trait loci and candidate genes involved in yield components under multiple field environments in cotton (Gossypium hirsutum). BMC Plant Biol 21:250. http://doi:10.1186/s12870-021-03009-2

Figures

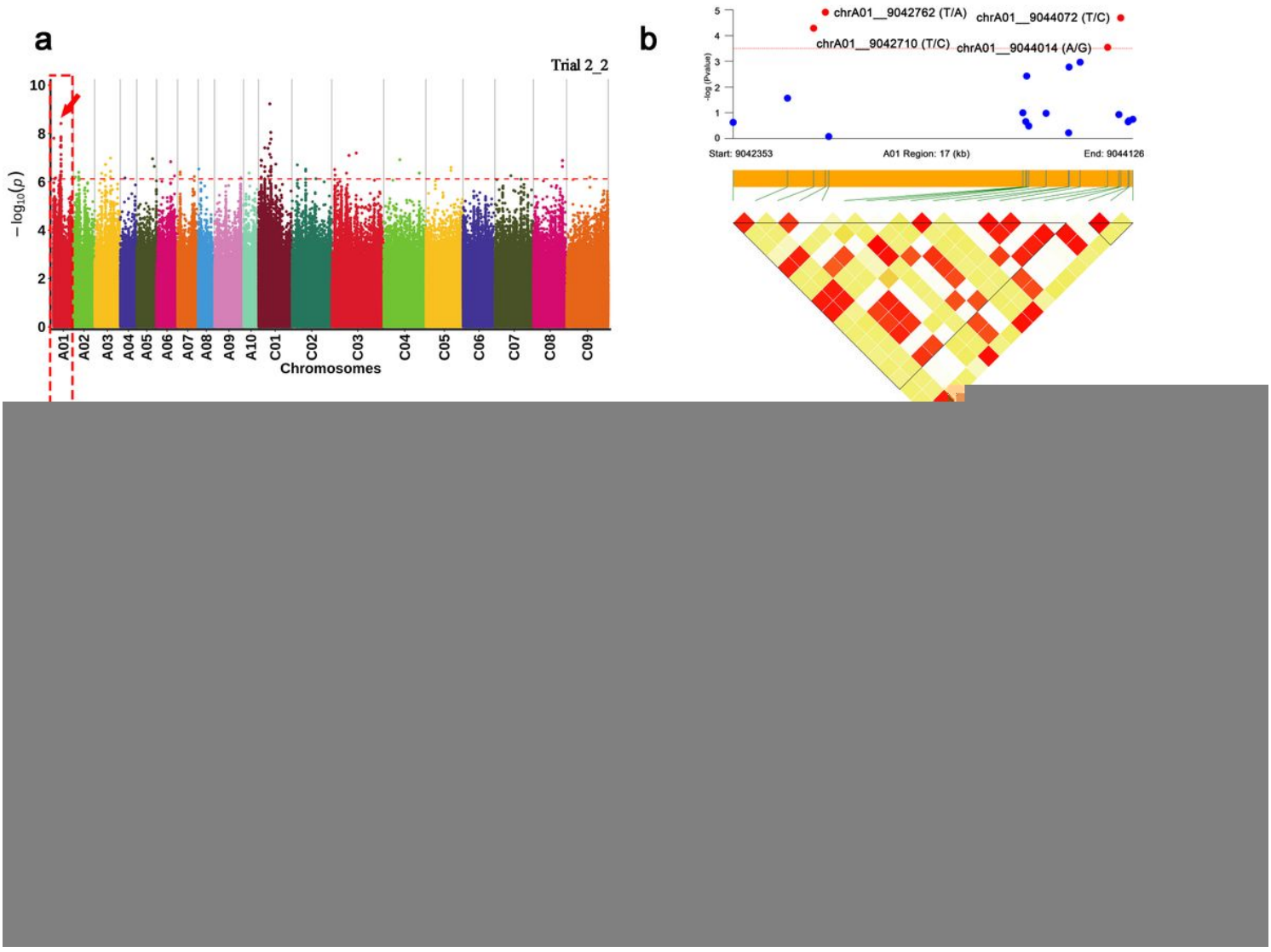


Figure 1

The co-localized locus and haplotypes on chromosome A01 associated with SY of *B. napus*. (a) Manhattan plot of co-localized locus for SY in Trial 2_2 and Trial 2_mean value. (b) Candidate gene association study of *BnaA01g17200D* with SY. Association of the four alleles in chrA01__9042710 (c), chrA01__9042762 (d), chrA01__9044014 (e), chrA01__9044072 (f) with SY, respectively. (g) Two haplotypes of *BnaA01g17200D*. The number of inbred lines harbouring the corresponding allele is shown in brackets at the bottom.

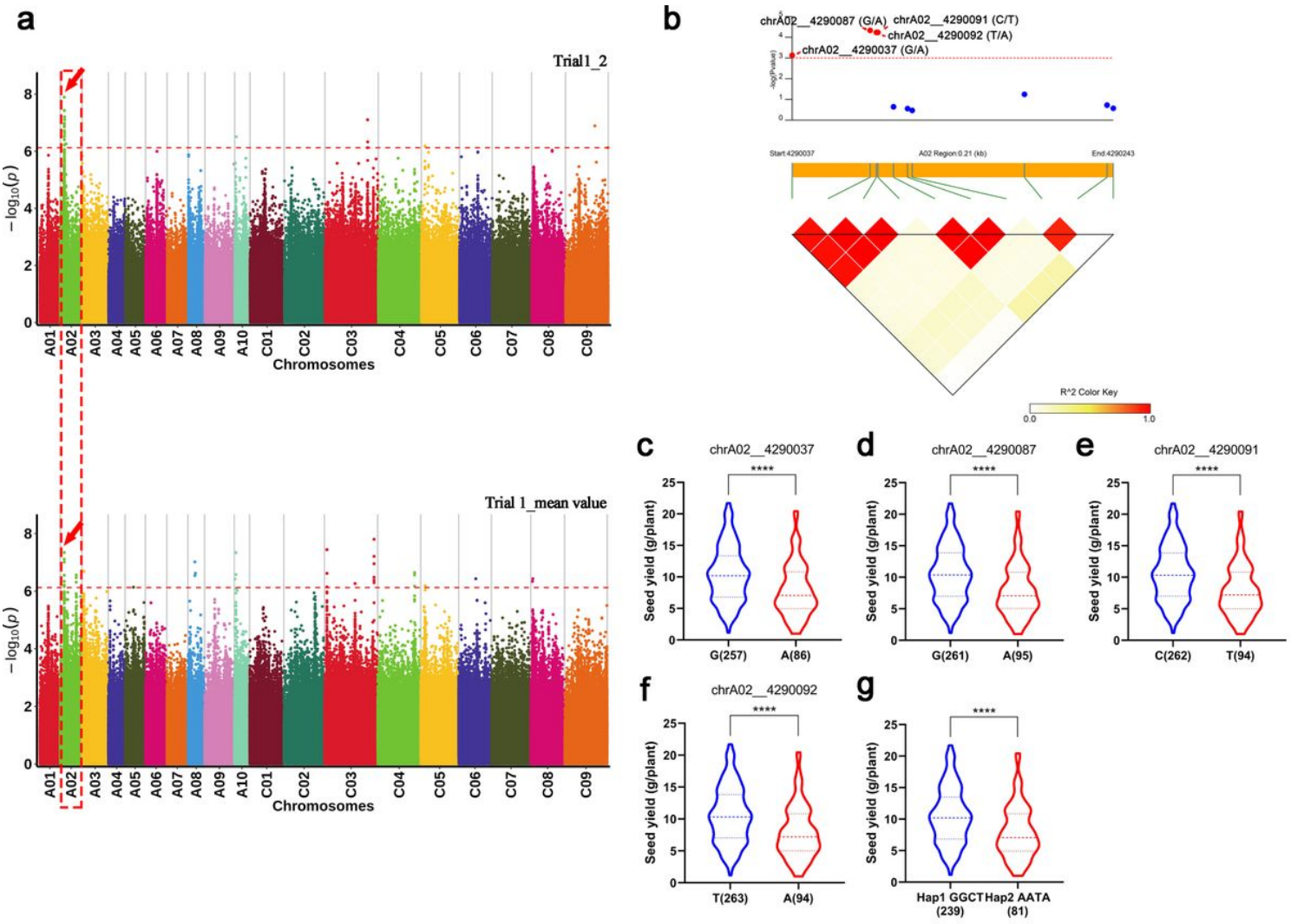


Figure 2

The co-localized locus and haplotypes on chromosome A02 associated with SY of *B. napus*. (a) Manhattan plot of co-localized locus for SY in Trial 1_2 and Trial 1_mean value. (b) Candidate gene association study of *BnaA02g08680D* with SY. Association of the four alleles in chrA02__4290037 (c), chrA02__4290087 (d), chrA02__4290091 (e), chrA02__4290092 (f) with SY, respectively. (g) Two haplotypes of *BnaA02g08680D*. The number of inbred lines harbouring the corresponding allele is shown in brackets at the bottom.

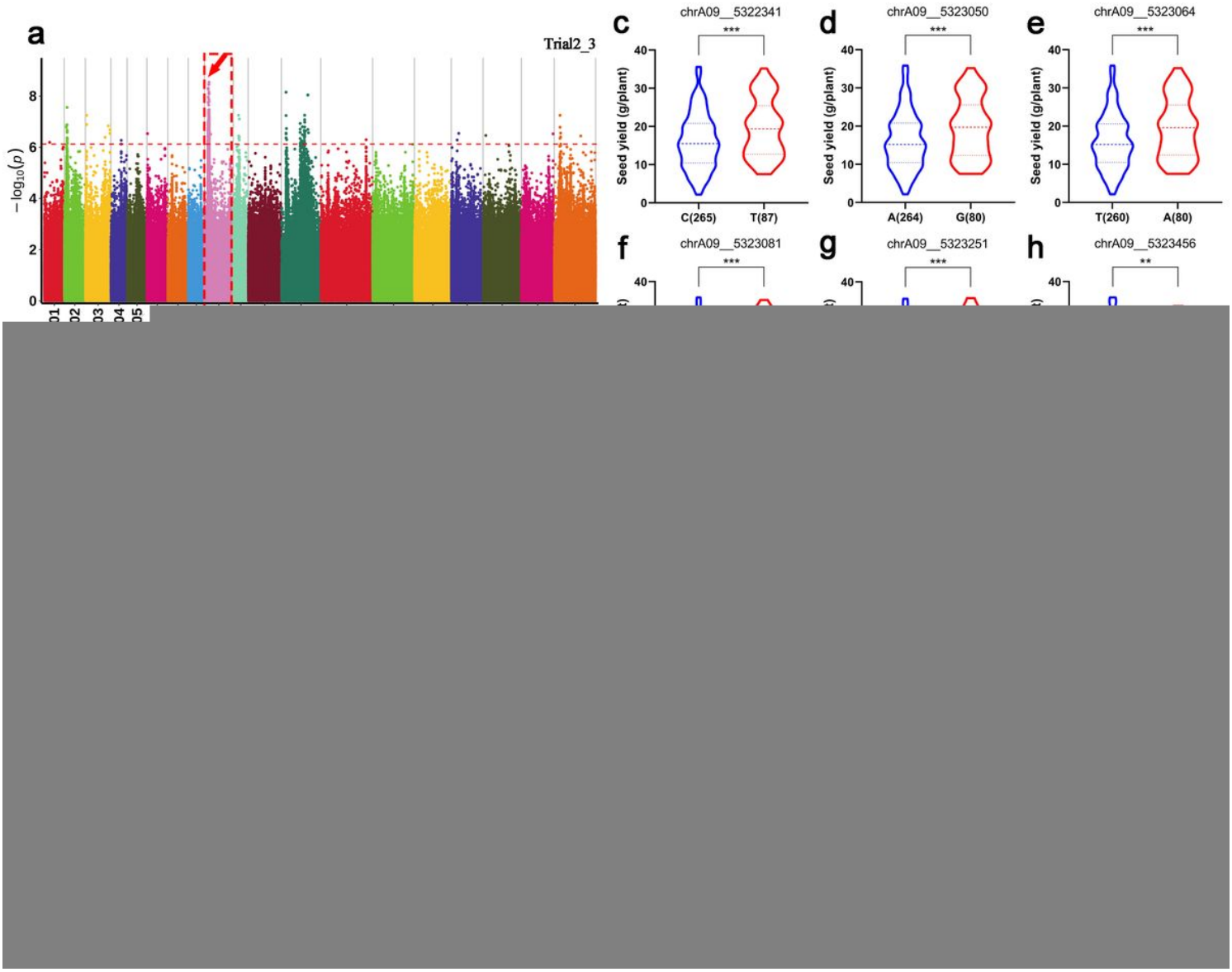


Figure 3

Candidate gene association study and haplotype analysis of *BnaA09g10430D*. (a) Manhattan plot of locus for SY in Trial 2_3. (b) Candidate gene association study of *BnaA09g10430D* with SY. Association of the eleven alleles in chrA09__5322341 (c), chrA09__5323050 (d), chrA09__5323064 (e), chrA09__5323081 (f), chrA09__5323251 (g), chrA09__5323456 (h), chrA09__5323496 (i), chrA09__5323633 (j), chrA09__5323759 (k), chrA09__5324271 (l), chrA09__5324376 (m) with SY, respectively. (n) Two haplotypes of *BnaA09g10430D*. The number of inbred lines harbouring the corresponding allele is shown in brackets at the bottom.

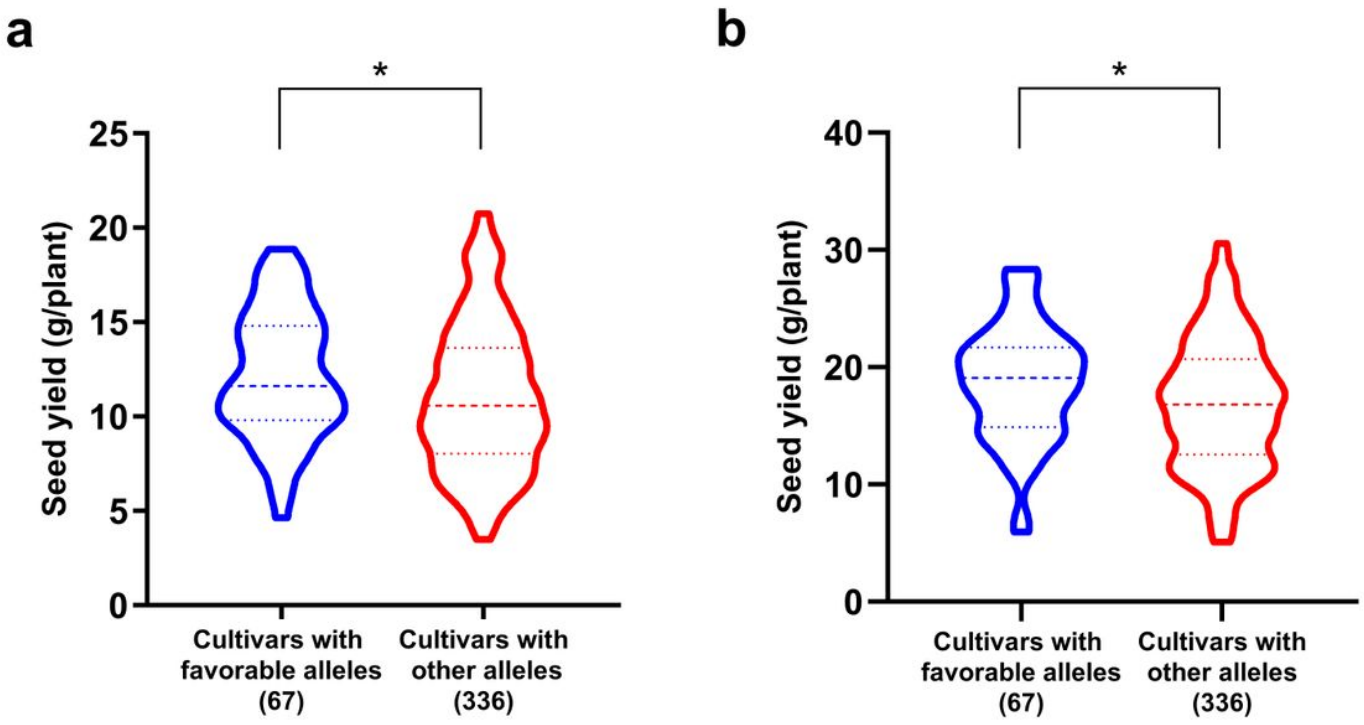


Figure 4

Differences in seed yield between *B. napus* varieties with three favorable haplotypes and other *B. napus* cultivars in the association panel. (a) Differences in Trial 1. (b) Differences in Trial 2.

Supplementary Files

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

- Fig.S1.docx
- Fig.S2.docx
- Fig.S3.docx
- Fig.S4.docx
- TableS1S10.xlsx