OXFORD

Full Paper

Earliness traits in rapeseed (*Brassica napus*): SNP loci and candidate genes identified by genome-wide association analysis

Qinghong Zhou[1], Depeng Han[1], Annaliese S. Mason[2], Can Zhou[1], Wei Zheng[3], Yazhen Li[3], Caijun Wu[1], Donghui Fu[1,]*, and Yingjin Huang[1,]*

[1]Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Agronomy College, Jiangxi Agricultural University, Nanchang 330045, China, [2]Plant Breeding Department, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen 35392, Germany, and [3]Jiangxi Institute of Red Soil, Jinxian, 331717, China

*To whom correspondence should be addressed. Tel. +086-0791-83828188. Fax. +086-0791-83828188. Email: fudhui@163.com (D.F.); yjhuang_cn@126.com (Y.H.)

Edited by Dr Kazuo Shinozaki

Abstract

Life cycle timing is critical for yield and productivity of *Brassica napus* (rapeseed) cultivars grown in different environments. To facilitate breeding for earliness traits in rapeseed, SNP loci and underlying candidate genes associated with the timing of initial flowering, maturity and final flowering, as well as flowering period (FP) were investigated in two environments in a diversity panel comprising 300 *B. napus* inbred lines. Genome-wide association studies (GWAS) using 201,817 SNP markers previously developed from SLAF-seq (specific locus amplified fragment sequencing) revealed a total of 131 SNPs strongly linked ($P < 4.96\text{E-}07$) to the investigated traits. Of these 131 SNPs, 40 fell into confidence intervals or were physically adjacent to previously published flowering time QTL or SNPs. Phenotypic effect analysis detected 35 elite allelic variants for early maturing, and 90 for long FP. Candidate genes present in the same linkage disequilibrium blocks ($r^2 > 0.6$) or in 100 kb regions around significant trait-associated SNPs were screened, revealing 57 *B. napus* genes (33 SNPs) orthologous to 39 *Arabidopsis thaliana* flowering time genes. These results support the practical and scientific value of novel large-scale SNP data generation in uncovering the genetic control of agronomic traits in *B. napus*, and also provide a theoretical basis for molecular marker-assisted selection of earliness breeding in rapeseed.

Key words: *Brassica napus*, genome-wide association study, SNP loci, candidate gene, earliness

1. Introduction

Flowering is a crucial transition from the vegetative growth phase to the reproductive growth phase in the plant life cycle, and flowering time is the most important factor in maintaining seed propagation in crop rotation systems. *Brassica napus* (AACC, $2n = 38$; rapeseed) is a young amphidiploid species (<10,000 years old) resulting from hybridization between *Brassica rapa* (AA, $2n = 20$; Asian cabbage, turnip) and *Brassica oleracea* (CC, $2n = 18$; European cabbage).[1,2] Rapeseed is the world's second largest oilseed crop, with an extensive economic impact in many countries worldwide.[3] Rapeseed oil is used not only for human consumption but also as industrial oil for lubricants and as biodiesel.[4] In southern China, farmers usually triple-crop annually: rice–rice–oil. With the introduction of *B. napus* into China in 1960, high-yielding *B. napus* varieties with disease-resistance and extensive adaptability outcompeted traditional *B. rapa* and *B. juncea* varieties, and are now planted widely in China.[5–7] However, due to the extended growth period of rapeseed, problems with using rapeseed in the existing crop rotation system are becoming more and more serious. The best way to solve this problem is to breed early-maturing varieties of *B. napus*, and thus early flowering and maturity have become major breeding goals in subtropical southern China. In temperate regions worldwide (e.g. Canada, the United States and Australia), early flowering and maturity are also important breeding targets for spring *B. napus* due to the short growth season.[8] In addition, flowering period (FP) is critical for *B. napus* production: longer FPs decrease the risk in hybrid seed production, as well as meeting travel market needs as a sightseeing attraction.[9] On the other hand, short and intensive FPs result in better uniformity of maturity time (MT), which is beneficial for mechanized harvesting of rapeseed.

Through extensive studies of genetic control networks for flowering time in *Arabidopsis thaliana*, we know that several pathways [vernalization, photoperiod, gibberellic acid (GA), autonomous pathway and thermal clock] and more than 100 genes are involved in the flowering process.[10] However, it is a difficult task to find the optimal time switch for flowering time in this gene network.[11–13] In the last few decades, QTL analysis has been used to investigate genetic variation for many complex agronomic traits in crops, and in particular flowering time traits. Many QTLs for flowering time in *B. napus* have been identified using bi-parental mapping populations.[14–22] However, bi-parental QTL analysis usually has the limitation of low resolution as well as low generalizability to crop breeding due to the participation of only two alleles from parents in the linkage mapping. In addition, some QTLs with small effects will not be detected, so these deficiencies need to be solved by other, newer methods.

Genome-wide association studies (GWAS), also known as association mapping or linkage disequilibrium (LD) mapping, aim to identify genetic variants linked to traits based on LD.[23] GWAS has the advantages of higher resolution and greater cost-effectiveness relative to bi-parental segregating populations, and excavates QTLs or genes from natural populations. GWAS performed with numerous SNPs has been used in *A. thaliana*, *Zea mays* and *Oryza sativa*.[24–26] In recent years, using the Illumina Infinium *Brassica* 60K SNP array, many studies used GWAS to detect genetic variation for agronomic traits in rapeseed. For example, Li et al. investigated the genetic architecture of seed weight and seed quality, detecting many significant marker–trait associations.[27] Liu et al. identified 50 loci significantly associated with seed oil content in a panel of 521 *B. napus* accessions.[28] Xu et al. explored 41 SNPs significantly correlated with flowering time, 12 SNPs of which were consistent with previously identified QTLs based on linkage analysis, and 25 candidate genes were implicated.[29] Schiessl et al. identified 101 genomic regions associated with initiation of flowering in 158 European winter-type *B. napus* inbred lines.[30] Wang et al. performed a genome-wide association analysis of flowering time with a panel of 448 rapeseed inbred lines, and found at least 40 QTLs significantly associated with flowering time: 117 genomic regions were found related to divergent growth habits, including 20 flowering time QTLs and 224 flowering time genes.[31] However, these research studies focused mainly on the trait of initiation of flowering time. Other traits related to flowering time, such as final flowering stage (FFS), FP and MT, have yet to be investigated via GWAS, and the identification of alleles conferring early flowering and maturity in *B. napus* have not been reported previously.

High-density SNP markers distributed across the whole genome are a prerequisite for genome-wide association analysis. Nowadays, it is possible to quickly and efficiently identify a large number of SNPs in a species via high-throughput DNA sequencing technologies.[32] In this study, 201,817 SNPs previously developed by SLAF-seq (specific length amplified fragment sequencing)[33] were used to perform a GWAS of four traits (IFS, FFS, FP and MT) in 300 inbred rapeseed lines. Correlations between these four traits were studied, SNP loci significantly associated with these traits and flowering time candidate genes were explored and elite allelic variants for earliness were identified. This study provides comprehensive information for understanding the relationship between flowering time variation and earliness traits. These SNPs and candidate genes detected will play an important role in earliness breeding in rapeseed.

2. Materials and Methods

2.1. Plant materials, growth conditions and field trials

A diversity panel consisting of 300 rapeseed inbred lines (S4 generation or greater) was used for the experimental population in the present study.[33] Pertinent information for all accessions is listed in Supplementary Table S1 (the population comprised 257 semi-winter types, 16 spring types and 27 winter types). The association population was grown in the field of Jiangxi Agricultural University (115.84E, 28.77N) and Jiangxi Institute of Red Soil (116.27E, 28.37N) with two replications per location in 2014–15 (designated JXAU and JXIRS, respectively), and all seeds were sown on 29 September 2014 simultaneously in both places. Each variety was planted in a plot with three rows (40 cm line width and 20 cm plant distance), and each row had 12 plants (final seeding time was at the 5–7 leaf phase). Field experiments were arranged by a randomized complete block design. Agronomic practices were kept uniform in both environments.

2.2. Phenotypic trait evaluation and statistical analysis

Dates for each of the four traits were recorded in the field trials: the initial flowering stage (IFS) (the number of days from sowing to the date when the first flower had opened in 25% of the plants in each plot), the FFS (the number of days from sowing to the date when 75% of the plants had stopped blooming completely in each plot), the FP (the number of days equal to the difference between the FFS and the IFS) and the MT (the number of days from sowing to the date when pods on 75% of the plants in each plot were yellow). The traits of each accession were defined as the mean of the two replicates in the same location. The correlation coefficients between traits were

determined using Student's t-test,[34] and the variance and statistical analysis of components were obtained using DPS software.[35]

Broad sense heritability was calculated as:

$$H_B^2 = \sigma_g^2 / \left(\sigma_g^2 + \sigma_{ge}^2 /_n + \sigma^2 /_{nr} \right),$$

where σ_g2 is the genetic variance, $\sigma_{ge}2$ is the variance term for the interaction between genotypes and environments, $\sigma2$ is the error variance, n is the number of environments and r is the number of replications in each experiment. This calculation is similar to Shi et al.[36]

2.3. SNP genotyping

Genomic DNA was extracted from young healthy leaves of each rapeseed accession using a modified cetyltrimethylammonium bromide method.[37] Quantified DNA was used for SLAF sequencing by an Illumina HiseqTM 2500.[38] Previously, through a set of processes of restriction digestion, library construction, paired-end sequencing and SNP calling, a series of 201,817 high-consistent and locus-specific SNPs (minor allele frequency > 0.05 and integrity > 0.8) were selected and used for subsequent analysis of population structure, LD and haplotype blocks in this diversity panel.[33]

2.4. Genome-wide association analysis

Based on the 201,817 SNP markers developed for the 300 rapeseed accessions, genome-wide association analysis for the four traits was carried out using general linear models (GLM) and mixed linear models (MLM) using the Tassel 5.0 software.[39] Fixed effects were calculated with a Q (population structure) matrix, and random effects were calculated with a K (Kinship) matrix. While only the Q matrix was taken into account in the GLM model, the Q + K matrices were both considered in the MLM model. The Q matrix was calculated using the Admixture software package,[40] and the K matrix (the genetic relationship among 300 accessions) was predicted using the SPAGeDi software.[41] P values for SNPs linked to traits were calculated using the following formula:

$$Y = X_\alpha + Q_\beta + K_\mu + e,$$

where Y represents the phenotype, X is the genotype, Q_β means fixed effect and K_μ means random effect. The Quantile–Quantile plot (Q-Q plot) was drawn by the GGplot2 software,[42] and the Manhattan plot was drawn by QQman software.[43] The threshold value of $-\log_{10} (P)$ was set as $-\log_{10} 0.1/201,817$ SNP [$P < 4.96E10\text{-}7$, $-\log_{10} (P)$ value is approximately equal to 6.3] for identifying true marker–trait associations, which was expressed as the false discovery rate (FDR) test value in the R program[44]; a true marker–trait association should show a FDR of less than 0.05, and only an FDR of less than 0.01 could meet the criteria for extremely significant association with the traits [$P < 4.96E10\text{-}8$, $-\log_{10} (P)$ value is approximately equal to 7.3].

2.5. Discovery of favourable allelic variation for earliness

For each of the trait-associated SNP loci, the phenotypic effect of each allelic variant was evaluated using the EAM method.[45] In addition, a trait has more than one associated SNP, so when the effect value is positive, we set it as increasing effect allele, and when the effect value is negative as a decreasing effect allele. The average allelic effect (AAE) was calculated with formula:

$$AAE = \sum a_c / n_c,$$

where a_c is effect value of c increasing or decreasing allele in a special SNP and n_c is the number of increasing or decreasing alleles.

SNPs with negative allelic effect values highly associated with IFS, FFS and MT were set as favourable alleles for earliness; SNPs with positive allelic effect values for prolonging FP were set as favourable alleles. The number of the favourable alleles in each rapeseed accession was counted.

2.6. Candidate genes for flowering time prediction

Based on LD analysis for the 300 accessions of rapeseed in our previous study,[33] the LD blocks where the significant trait-associated SNPs were situated, in which flanking SNP markers had strong LD ($r^2 > 0.6$),[46] were defined as the candidate gene regions (extending from the left unrelated SNP to the right unrelated SNP). The LD block was analysed using the software 'haploview v4.2'.[47] All genes within the same LD block ($r^2 > 0.6$) as significantly trait-associated SNP markers were considered for identification of candidates. For significant SNPs outside of the LD blocks, the 100 kb flanking regions on either side of the markers were used to identify candidate genes. All candidate genes were selected based on gene ontology (GO) terms for flowering, floral development, vernalization, photoperiod and vegetative to reproductive phase.[30,33] Subsequently, we carried out BLASTX searches against the *Arabidopsis* genome to determine the final flowering time candidate genes within the SNP-tagged genome regions.

2.7. Comparison of SNPs and QTLs related to flowering time traits

A genomic region of 200 kb (roughly equal to 0.4 cM in genetic map) was set as a single QTL identified in previous research.[46,48] These QTL regions containing trait-linked SNPs were compared with the results of our study, and to GWAS results detecting flowering time gene loci using the Illumina Infinium *Brassica* 60K SNP array to map SNPs to physical positions in the *B. napus* genome.[49] In addition, by anchoring known marker sequences (SSR, RFLP, etc.) to the rapeseed reference genome within the range of 1 Mb, SNPs previously connected to flowering time and MT QTLs in bi-parental mapping populations, as collected in our published article,[50] were also compared with the results of our study.

3. Results

3.1. Phenotypic variation and correlation analysis for the four earliness traits in 300 rapeseed accessions

Four traits related to earliness of 300 rapeseed lines were investigated in two environments in this study. Table 1 showed that the average time to IFS was 145.07 and 150.63 days with the coefficient of variation of 9.88% and 10.11%, respectively, in environments JXAU and JXRIS (Table 1); the minimum was 91 days and the maximum was 192 days. Analogously, the FFS also exhibited a wide range of 169–208 and 163–215 days, with means of 180 and 186 days in JXAU and JXRIS, respectively. The mean value of FP in JXAU was 35.57 days, ranging from 12 to 83 days with a coefficient of variation of 29.21%, and the average number of days for FP in JXIRS was 36.54 with a coefficient of variation of 31.06% (varying from 17 to 97 days), such that large variation was clearly observed for these traits. Finally, the average number of days to maturity was

Table 1. Statistical analysis of earliness traits of rapeseed in two environments (JXAU and JXIRS)

Environment	Trait	Mean ± SD (d)	Mode	Min/d	50% Quantile/d	Max/d	CV (%)	Shapiro–Wilk test
JXAU	IFS	145.07±14.33	142	91	143	186	9.88	W=0.912653 P=0.000001
	FFS	180±5.98	176	169	180	208	3.31	W=0.826885 P=0.000001
	FP	35.57±10.39	36.5	12	36	83	29.21	W=0.876773 P=0.000001
	MT	217.41±4.90	217.5	203	218	234	2.25	W=0.978996 P=0.000218
JXIRS	IFS	150.63±15.23	151	91	151.5	192	10.11	W=0.959349 P=0.000001
	FFS	186.94±7.64	185.5	163	187	215	4.08	W=0.932065 P=0.000001
	FP	36.54±11.35	34.5	17	34	97	31.06	W=0.845379 P=0.000001
	MT	219.80±4.03	218.5	210	219	235	1.83	W=0.913101 P=0.000001

CV: Coefficient of variation.

217.41 days in JXAU, ranging from 203 to 234 days with a coefficient of variation of 2.25%, and the time to maturity in JXIRS was 219.80 days ranging from 210 to 235 days, which exhibited the lowest coefficient of variation of 1.83%. Extensive variation for each of the four traits was observed in two environments, and phenotypic values for each of the four traits were normally distributed (Fig. 1). In addition, the average days to IFS and FFS in environment JXAU was later than in environment JXIRS, by about 5 and 6 days, respectively. However, the FP was generally consistent between the two environments, and the MT in the JXAU environment was only earlier than in the JXIRS environment by 2 days on average. These data implied a broad diversity in earliness phenotypic traits in the population of 300 rapeseed accessions.

Analysis of variance (ANOVA) was conducted for the 300 accessions to test the effects of genotype (G), environment (E) and their interactions (G × E) for the four traits. All traits varied significantly across the 300 genotypes ($P < 0.01$; Supplementary Table S2), and there were obvious differences in FP between the two environments ($P < 0.05$), as well as in the other three traits between the two environments ($P < 0.01$). However, differences in traits between repetitions were not significant, although G × E interactions were all significant ($P < 0.01$), suggesting a large environmental impact on these traits in rapeseed. The broad-sense heritability of IFS was calculated to be 95.42%, while FFS, FP and MT had broad sense heritabilities of 92.35%, 91.99% and 87.55%, respectively. All traits were stably inherited with an H_B^2 higher than 85%.[36]

Initial flowering in the two environments had a highly significant positive correlation with FFS and MT (Table 2), with phenotypic correlation coefficients of 0.7774** and 0.5698** in JXAU and 0.7053** and 0.5118** in JXIRS, respectively, indicating that early flowering means early maturity, with flowering time a crucial indicator for MT. However, FP had a highly significant negative correlation with the other three traits, with phenotypic correlation coefficients of −0.932**, −0.4976** and −0.4053** in JXAU and −08757**, −0.2772** and −0.2903** in JXIRS, respectively, indicating that the sooner flowering time and MT are reached, the longer the FP is, and vice versa.

3.2. Genome-wide association analysis for the four earliness traits in the 300 rapeseed accessions

To uncover the genotypic variations of four traits related to earliness in *B. napus*, GLM and MLM models for GWAS were evaluated, and the degree of consistency between the observed and expected *P* values were assessed using QQ plots, both models controlled the generation of the false positives well (Fig. 2), and the significant SNPs associated with traits were displayed on Manhattan plots (Table 3,

Figs 3 and 4). The GLM analysis detected a total of 125 SNPs ($P < 4.96E-07$) significantly associated with four earliness-related traits, and distributed on 18 of the 19 *B. napus* chromosomes (excluding A04). The largest number of significant SNPs (25) was on chromosome C01, and the second largest number (18 SNPs) was found on chromosome C03 (Fig. 4, Table 3). MLM analysis detected 22 SNPs significantly associated with IFS (3), FFS (3) and FP (18) on 11 chromosomes, 2 of which were associated with IFS and FP simultaneously on chromosome A09 (Table 3). Totally, 131 SNPs significantly associated with four traits were detected on 18 chromosomes by both GLM and MLM analyses.

Specifically, 26 SNPs were associated with initial flowering in the two environments, 9 of which were detected in both environments by GLM model analysis (Supplementary Table S2). However, only three SNPs for initial flowering time were found in environment JXAU by MLM analysis, all on chromosome A09 (Supplementary Table S4). Using GLM analysis, 10 SNPs associated with FFS were detected (8 in JXAU and 2 in JXRIS) but no consistent SNPs for FFS were detected in both environments. Using MLM analysis, three SNPs on chromosome A01 in environment JXAU for FFS were detected, one of which was consistent with the GLM predictions (Supplementary Tables S5 and S6). The GLM model detected 106 SNPs associated with FP in environment JXAU, of which 49 SNPs were extremely significant ($P < 4.96E-08$; Supplementary Table S7), while 78 SNPs associated with FP were detected in environment JXIRS, 22 of which were extremely significant; 16 SNPs were identified in both environments (Supplementary Table S8). MLM analysis identified 18 SNPs for FP (4 in JXAU and 15 in JXIRS with one shared SNP locus in two environments), 15 SNPs of which were consistent with GLM analysis. However, for MT, only nine SNPs were detected in one environment (JXAU) using the GLM model (Supplementary Table S9).

3.3. Discovery of useful allelic variation for earliness

In order to identify elite alleles for earliness breeding in *B. napus*, we evaluated the allelic effects of SNP loci associated with four earliness-related traits. SNP alleles with positive effects that led to decreases in trait values for IFS, FFS and MT, or that led to an increase in the trait value for FP, were defined as 'favourable alleles' for earliness. In our study, we observed 35 favourable alleles from 29 SNP loci for earliness of flowering and maturity, which were present in 288 accessions. Individual accessions had from 1 to 17 favourable alleles for earliness, with 19 accessions having more than 10 alleles for earliness. For this latter group of accessions, mean IFS (130 days) was shorter by 15 days than the mean for all 300 accessions (Supplementary Tables S10 and S11). Figure 5a shows that more favourable alleles resulted in earlier flowering or maturity. For the
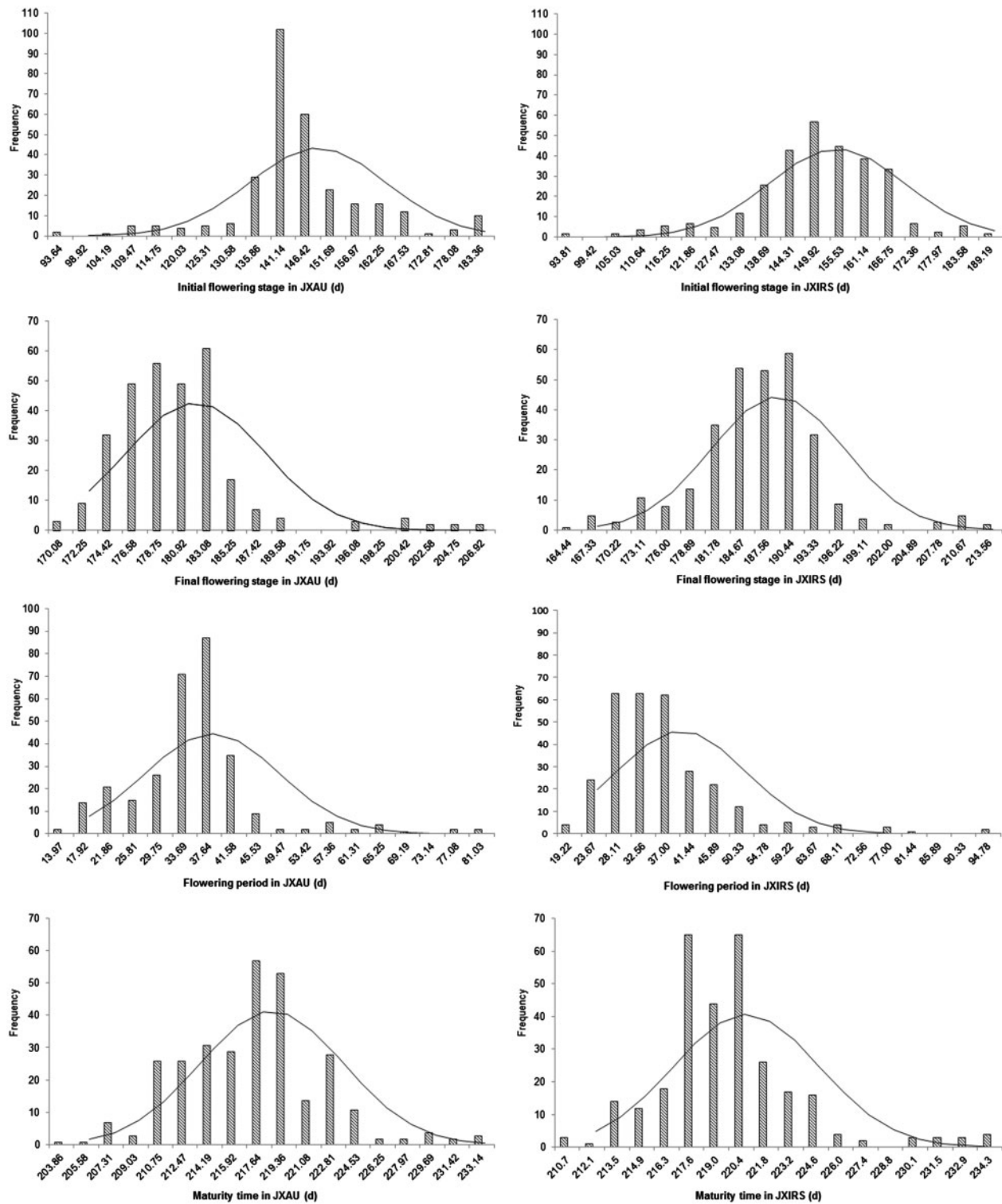
Figure 1. Frequency distribution of four traits related to earliness of *Brassica napus* in two environments (JXAU and JXIRS).
Note: The *X*-axis indicates the trait (days) and *Y*-axis indicates the accession number.

FP, based on the assessment of allelic effect values, 74 SNP loci (90 alleles) contributed to a prolonged FP, with the number of favourable alleles per accession ranging from 5 to 47. A set of 13 rapeseed lines had more than 20 alleles for long FP, and this group had a much longer average FP (56 days) than the average across all

accessions (35 days) (Supplementary Table S12; Fig. 5b shows a long FP phenotype). Furthermore, by comparing the allelic effect values of alleles between the traits of IFS and FP, all alleles promoting early flowering were found to be totally consistent with prolonged FP in all accessions (Supplementary Tables S10, S11 and S13). Among

Table 2. Correlation analyses of earliness traits of rapeseed in two environments (JXAU and JXIRS)

Correlation	IFS/d	FFS/d	FP/d	MT/d
IFS/d	1			
FFS/d	0.7774**/0.7053**	1		
FP/d	−0.932**/-0.8757**	−0.4976**/-0.2772**	1	
MT/d	0.5698**/0.5118**	0.6605**/0.6005**	−0.4053**/-0.2903**	1

* and ** represents Significance at 5% ($P = 0.1133$) and 1% ($P = 0.1485$) probability levels, respectively.
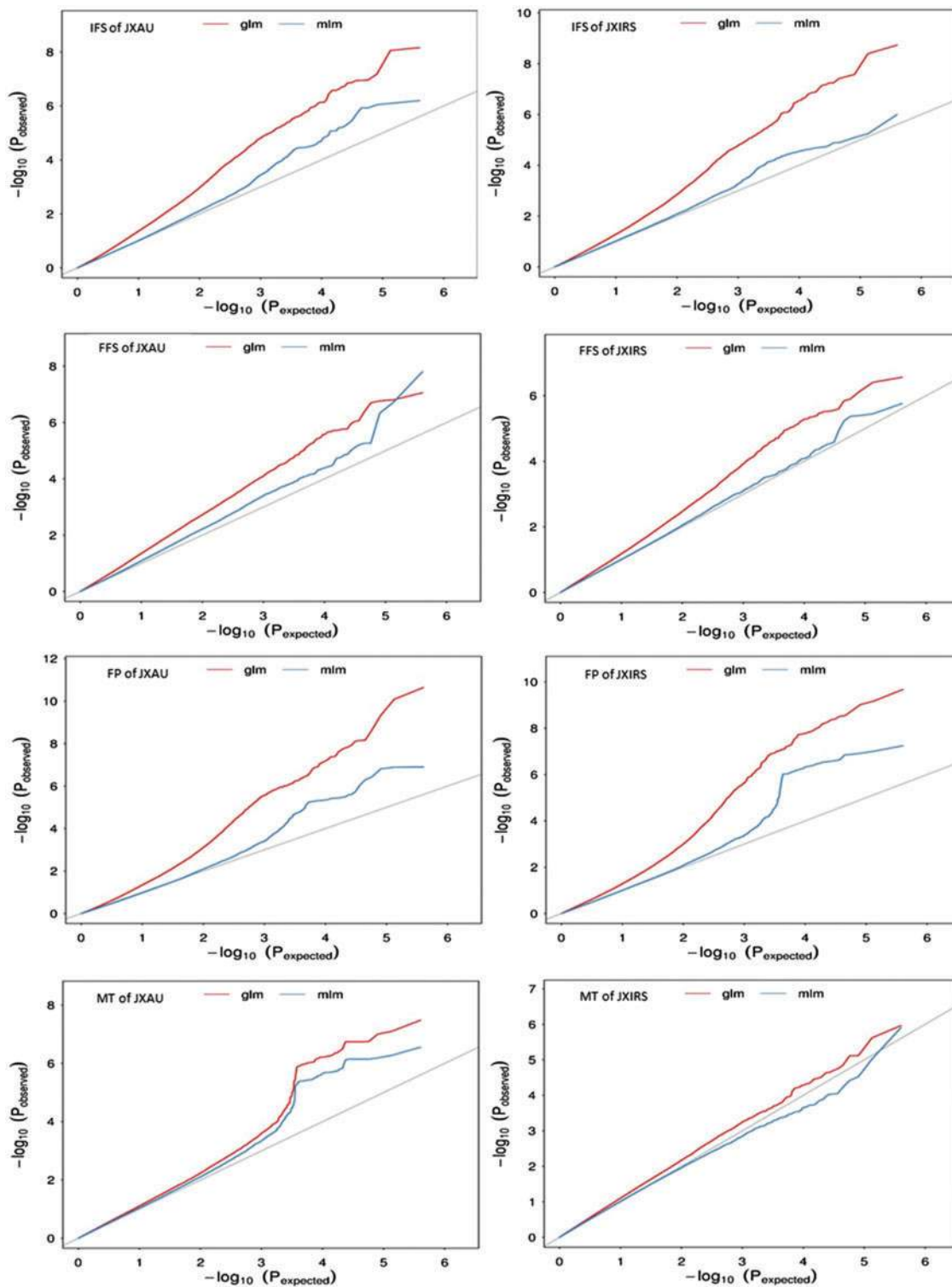

Figure 2. Quantile-quantile plots for four traits related to earliness using two models with GLM (upper curve) and MLM (lower curve) in two environments (JXAU and JXIRS).
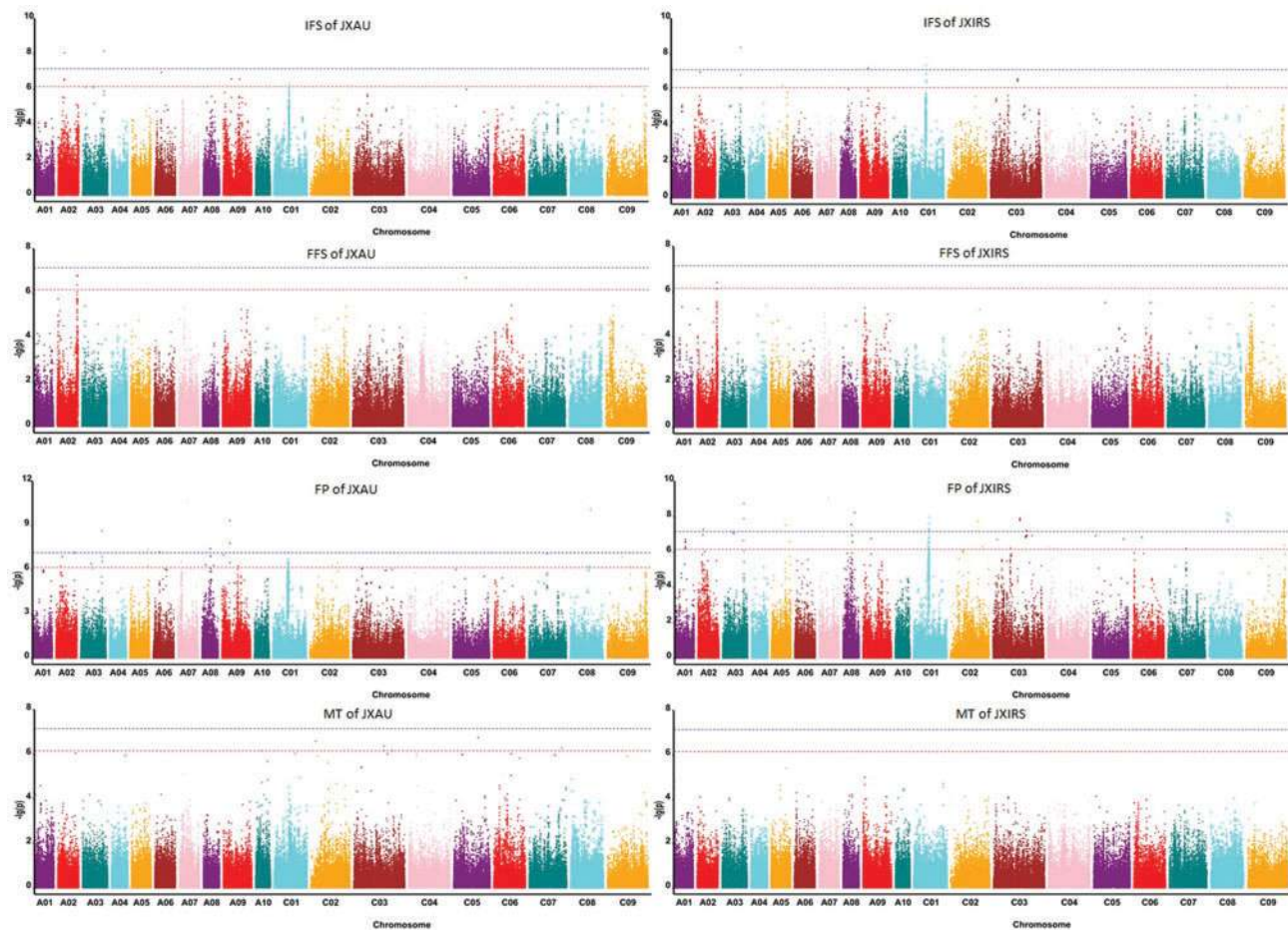
Figure 3. Manhattan plot for four traits related to earliness of *Brassica napus* by GLM model in two environments (JXAU and JXIRS).
Note: The lower dashed horizontal line represents the significance threshold ($P < 4.96E10-7$, $-\log10 (P)$ value is approximately equal to 6.3); the upper dashed horizontal line represents the extreme significance threshold ($P < 4.96E10-8$, $-\log10 (P)$ value is approximately equal to 7.3).

these trait-linked loci, we detected 6 SNPs with 2 favourable allelic variations for earliness in flowering or maturity on chromosomes A01, A02, A09, C03 and C04, and another 16 SNPs with favourable alleles for prolonging FP were distributed on chromosomes A05, A06, A08, C01, C02, C03, C08 and C09. In addition, some SNP loci had different effects associated with each of the two alleles present at the SNP locus: for example, the favourable C allele of *Bn-A02-23681432* and *Bn-A02-23875175* related to FFS was associated with earlier MT compared with the unfavourable G allele, with an average of about 3 days of phenotypic difference in the two environments (see Fig. 6). These results indicate that the highly favourable SNP alleles exhibit significant positive effects on phenotypic characteristics compared with the unfavourable alleles.

3.4 Identification of candidate genes for flowering time in *B. napus*

Of the 131 SNP loci significantly associated with earliness traits, 85 SNP loci were divided into 29 candidate genome regions based on the LD blocks analysis ($r^2 > 0.6$), ranging in size from 253 bp to 576.661 kb (Supplementary Table S14), while the remaining 46 SNP loci were not present in the defined LD blocks. To further uncover the molecular function of the significant SNPs, we obtained the genes within the same LD block or within 100 kb to either side of the

significant trait-associated SNPs by BLAST analysis using *B. napus* 'Darmor v4.1' as the reference genome. We screened 1,672 genes in the candidate regions of 80 SNPs significantly associated with the four earliness traits (Supplementary Table S15): 147 candidate genes closely linked with 44 SNPs were obtained based on the GO terms related to flowering time (flowering, floral development, vernalization, photoperiod, vegetative to reproductive transition and gibberellin) (Supplementary Table S16). Of these, 57 flowering time candidate genes closely linked with 33 SNPs in *B. napus* were identified as orthologous to *A. thaliana* genes in flowering time networks, which were involved in the flowering regulation pathways of vernalization, photoperiod, GA, autonomous pathway and circadian clock, respectively (Supplementary Figure S2).[10] These candidate genes were distributed on 14 chromosomes, 44 of which were distributed on the A subgenome, with the most genes (12) on chromosome A02 and with the other 13 flowering genes located on 5 chromosomes of the C subgenome (Supplementary Table S17, Fig. 7).

In the vicinity of some SNP regions, more than one known flowering time gene was identified (Fig. 7). For example, at the position of SNP *Bn-A03-16342394*, we found five flowering time candidate genes (orthologous to *A. thaliana* genes of *ABF4*, *PIE1*, *VRN1* and *ARP6*), at the position of SNP *Bn-A10-13390883* three important flowering time candidate genes (orthologous to *A. thaliana* genes of *CO*, *SVP* and
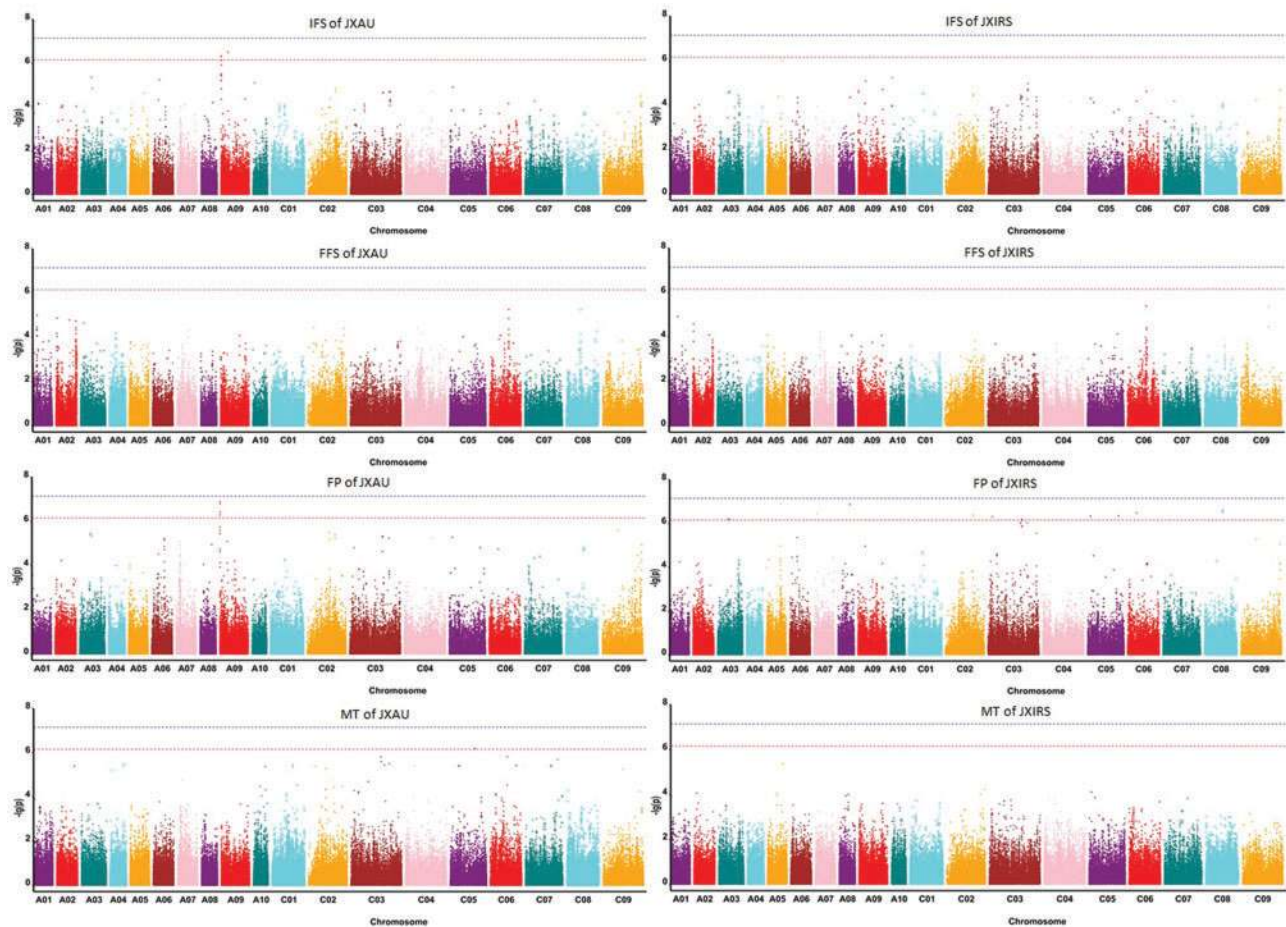
Figure 4. Manhattan plot of four traits related to earliness of *Brassica napus* by MLM model in two environments (JXAU and JXIRS).
Note: The lower dashed horizontal line represents the significance threshold ($P < 4.96E10-7$, $-\log10$ (*P*) value is approximately equal to 6.3), the upper dashed horizontal line represents the extreme significance threshold ($P < 4.96E10-8$, $-\log10$ (*P*) value is approximately equal to 7.3).

AtCOL1) were found, at the position of SNP *Bn-A03-13764115* four flowering time genes (*VGT1*, *IPP2*, *GRF7* and *COL2*) were found, at the position of SNP (*Bn-A02-23681432*) three flowering time genes (*CDF1*, *HUA2* and *CHE*) were detected, and three flowering time genes (*FT*, *RGL1* and *FLD*) were closely adjacent to the position of SNP *Bn-A02-6435246*. These results indicate that some loci are highly associated with flowering time genes, and that genes controlling flowering tend to be located in clusters in *B. napus*.[51] In addition, we found the significant SNP locus *Bn-A03-25129078* is in the inner region of the flowering time candidate gene of *GSBRNA2T00044315001* homologous to *CURLY LEAF* (*CLF*) of *A. thaliana*. By evaluating the allelic effect of *Bn-A03-25129078* locus, accessions with an A allele for *Bn-A03-25129078* had an average of 130 days for IFS in the two environments, 19 days earlier than accessions with the G allele. This indicates that *Bn-A03-25129078* is an important SNP locus in promoting early flowering in rapeseed, which also proves the reliability for identifying the candidate genes using GWAS analysis.

4. Discussion

4.1. Identification and validation of SNP loci associated with traits related to earliness in *B. napus*

Identifying favourable allelic variation and candidate genes promoting early flowering and maturity is critical for effective use of

B. napus in tri-annual crop rotation systems in China. Earliness of *B. napus* is a very important trait for reducing the planting time conflict during tri-annual crop rotation systems in southern China. However, earliness is a complicated quantitative trait. In previous studies, flowering time in *B. napus* showed a high genetic correlation (0.73) with MT,[17,22,52,53] with QTLs co-localized with plant height in a small region on chromosome A02.[20,54] By *in silico* QTL integration, co-localization of flowering time and MT was also identified on chromosomes A01, A02, A03, A05 and C09.[50] Therefore, flowering time is a crucial indicator for MT. In this study, we investigated four traits related to earliness (IFS, FFS, FP and MT) for 300 rapeseed accessions in two environments. High correlations between these traits were also identified, and strong positive correlations existed between the traits of IFS, FFS and MT. Furthermore, we found that the trait of FP was highly negatively correlated with the other three traits. By evaluating the allelic effects of SNP loci associated with four earliness related traits, we revealed that favourable alleles promoting early-flowering and early-maturing are totally opposite to the favourable alleles prolonging flowering days in all accessions. Therefore, we can infer that the rapeseed varieties with alleles for earliness should have longer FPs.

In the current study, all four traits showed large phenotypic variation in the two environments, supporting the suitability of genome-wide association analysis for these traits using this diversity panel.

Table 3. SNP loci significantly associated with earliness traits in *Brassica napus*

SNPs	Chromosome	Position	P value	R² (%)	Allele	Environment		Traits				Methods	
						JXAU	JXIRS	IFS	FFS	FP	MT	GLM	MLM
Bn-A02-p7537706	A02	7537706	1.28E-07*-1.63E-09**	8.26–9.34	G/T	○		√		√		●	
Bn-A02-p7272190	A02	7272190	4.79E-07*	5.86	A/G	○		√				●	
Bn-A02-p7272126	A02	7272126	2.63E-07*	5.91	A/G	○		√				●	
Bn-A03-p25129078	A03	25129078	1.54E-08**-1.83E-09**	8.41–13.07	G/A	○	○	√		√		●	
Bn-A06-p7621426	A06	7621426	6.26E-08*-3.75E-08**	6.51–9.21	C/T	○		√		√		●	
Bn-A07-p10864549	A07	10864549	2.08E-07*-2.29E-11**	6.39–12.39	A/T	○	○	√		√		●	
Bn-A09-p9520665	A09	9520665	1.67E-07*-3.72E-08**	6.51–9.11	A/T	○	○	√		√		●	
Bn-A09-p19488065	A09	19488065	7.15E-08*	5.15	T/C	○		√				●	
Bn-C01-p17761695	C01	17761695	4.70E-07*-1.05E-08**	6.25–11.83	T/G	○	○	√		√		●	
Bn-SA03-p2039163	scaffoldA03_random	2039163	1.12E-07*-2.12E-10**	6.2–12.79	A/G	○	○	√		√		●	
Bn-SA08-p1469091	scaffoldA08_random	1469091	2.90E-07*-2.17E-08**	5.8-9.07	C/T	○	○	√		√		●	
Bn-SA08-p1469110	scaffoldA08_random	1469110	2.37E-07*-1.77E-08**	5.79–9.17	T/C	○	○	√		√		●	
Bn-A02-p6435246	A02	6435246	9.34E-08*-5.08E-08*	8.27–10.67	A/G		○	√		√		●	
Bn-A03-p25169892	A03	25169892	1.40E-08*-1.31E-08**	8.07–11.34	A/G	○	○	√		√		●	
Bn-C01-p17761632	C01	17761632	1.26E-07*-5.57E-08*	8.24–10.74	A/G		○	√		√		●	
Bn-C01-p17761646	C01	17761646	3.37E-07*-9.15E-08*	7.69–10.36	A/C		○	√		√		●	
Bn-C01-p17761654	C01	17761654	2.75E-07*-5.55E-08*	7.85–10.74	C/T		○	√		√		●	
Bn-C01-p17761705	C01	17761705	3.40E-07*-5.11E-08*	7.96–10.93	A/G		○	√		√		●	
Bn-C03-p32042627	C03	32042627	2.71E-07*-1.43E-08**	8.74–12.81	A/C		○	√		√		●	
Bn-C03-p32042629	C03	32042629	3.08E-07*-1.65E-08**	8.73–12.76	A/G		○	√		√		●	
Bn-C03-p32042644	C03	32042644	2.29E-07*-1.24E-08**	9.03–13.17	C/T		○	√		√		●	
Bn-SA02-p297545	scaffoldA02_random	297545	7.89E-08*-2.67E-08**	9.30–10.97	A/G	○	○	√		√		●	
Bn-SA09-p3882263	scaffoldA09_random	3882263	4.88E-07*-3.07E-07*	7.13–9.09	A/C		○	√				●	
Bn-A09-p122597	A09	122597	4.91E-07*-3.72E-07*	9.14–10.31	A/C	○		√		√			●
Bn-A09-p122628	A09	122628	4.69E-07*-1.26E-07*	9.29–11.09	A/T	○		√		√			●
Bn-A09-p8430274	A09	8430274	3.15E-07*	8.6	C/T	○		√					●
Bn-A02-p23681151	A02	23681151	3.95E-07*	3.39	A/G	○			√			●	
Bn-A02-p23681432	A02	23681432	4.23E-07*	3.55	C/G	○			√			●	
Bn-A02-p23875175	A02	23875175	1.71E-07*	4.69	C/G	○			√			●	
Bn-A02-p23875300	A02	23875300	1.57E-07*	4.53	A/G	○			√			●	
Bn-C05-p15929590	C05	15929590	2.00E-07*	3.27	C/T	○			√			●	
Bn-SA01-p311119	scaffoldA01_random	311119	8.78E-08*-1.57E-08**	3.51–10.71	G/T	○			√			●	●
Bn-A02-p23924336	chrA02	23924336	3.92E-07*	6.1	C/T		○		√			●	
Bn-SC04-p1483308	scaffoldC04_random	1483308	2.74E-07*	6.09	G/T		○		√			●	
Bn-SA01-p311144	scaffoldA01_random	311144	1.94E-07*	9.15	G/T	○			√				●
Bn-SA01-p311434	scaffoldA01_random	311434	4.65E-07*	8.61	A/T	○			√				●
Bn-A02-p21988229	A02	21988229	7.10E-08*	9.25	G/G	○				√		●	
Bn-A03-p12287810	A03	12287810	3.94E-07*-7.97E-08*	7.91–9.93	G/T	○	○			√		●	
Bn-A03-p12288072	A03	12288072	3.94E-07*-7.97E-08*	7.91–9.93	A/G	○	○			√		●	
Bn-A03-p25225210	A03	25225210	2.80E-07*	8.07	G/T	○				√		●	
Bn-A05-p16342394	A05	16342394	2.68E-07*-4.13E-08**	9.71–9.86	C/T	○	○			√		●	
Bn-A05-p21493588	A05	21493588	3.05E-07*-4.13E-08**	9.71–9.86	C/A	○	○			√		●	
Bn-A06-p15620541	A06	15620541	2.14E-07*	7.14	A/T	○				√		●	
Bn-A07-p3952619	A07	3952619	3.25E-07*	8.77	A/G	○				√		●	
Bn-A07-p3986920	A07	3986920	2.57E-07*	9.55	A/G	○				√		●	
Bn-A07-p3986954	A07	3986954	8.17E-08*	9.57	A/C	○				√		●	
Bn-A08-p4107097	A08	4107097	4.79E-07*	9.29	A/G	○				√		●	
Bn-A08-p9750352	A08	9750352	2.79E-08**-3.72E-08**	9.11–10.41	A/T	○	○			√		●	
Bn-A08-p10241291	A08	10241291	1.23E-07*	8.29	G/T	○				√		●	
Bn-A08-p10241315	A08	10241315	1.23E-07*	8.29	C/T	○				√		●	
Bn-A09-p1622239	A09	1622239	9.86E-08*	9.84	A/T	○				√		●	
Bn-A09-p9517894	A09	9517894	3.27E-07*	7.83	A/T	○				√		●	
Bn-A09-p9526166	A09	9526166	1.61E-08**	9.34	C/G	○				√		●	
Bn-C01-p17478274	C01	17478274	1.79E-07*	8.03	A/C	○				√		●	
Bn-C01-p17478312	C01	17478312	3.36E-07*	7.69	C/T	○				√		●	
Bn-C01-p17726513	C01	17726513	6.67E-08*	8.6	C/T	○				√		●	
Bn-C01-p17726760	C01	17726760	3.59E-07*	7.77	C/G	○				√		●	
Bn-C01-p17726781	C01	17726781	4.58E-07*	7.65	C/G	○				√		●	
Bn-C01-p17726814	C01	17726814	3.59E-07*	7.77	C/G	○				√		●	

Continued

Table 3. continued

SNPs	Chromosome	Position	P value	R² (%)	Allele	Environment JXAU	Environment JXIRS	Traits IFS	Traits FFS	Traits FP	Traits MT	Methods GLM	Methods MLM
Bn-C01-p17759676	C01	17759676	4.22E-07*	7.67	C/G	○				√		●	
Bn-C01-p17759711	C01	17759711	4.22E-07*	7.67	G/T	○				√		●	
Bn-C01-p17797361	C01	17797361	4.87E-07*	7.73	C/T	○				√		●	
Bn-C02-p33057434	C02	33057434	3.47E-07*-1.85E-08**	7.73–10.56	C/T	○	○			√		●	●
Bn-C03-p20184518	C03	20184518	2.24E-07*	9.09	C/G	○				√		●	
Bn-C07-p21671359	C07	21671359	8.31E-08*	8.62	C/T	○				√		●	
Bn-C08-p23973542	C08	23973542	8.37E-09**-8.15E-11**	10.84–11.67	C/T	○	○			√		●	
Bn-C09-p18507822	C09	18507822	1.43E-07*	8.87	C/T	○				√		●	
Bn-A01-p12901579	A01	12901579	2.61E-07*	10.08	A/G		○			√		●	
Bn-A01-p12901581	A01	12901581	2.79E-07*	10.08	C/T		○			√		●	
Bn-A01-p12901588	A01	12901588	2.61E-07*	10.08	A/G		○			√		●	
Bn-A01-p12901611	A01	12901611	1.96E-07*	10.99	C/T		○			√		●	
Bn-A01-p12901881	A01	12901881	4.95E-07*	10.43	A/G		○			√		●	
Bn-A02-p5752414	A02	5752414	1.09E-07*	10.74	A/G		○			√		●	
Bn-A03-p13764115	A03	13764115	8.88E-08*	9.49	G/T		○			√		●	
Bn-A03-p24569600	A03	24569600	2.29E-07*	9.47	G/T		○			√		●	
Bn-A05-p17034333	A05	17034333	1.20E-07*-3.13E-08**	10.81–11.66	A/T		○			√		●	●
Bn-A05-p21798836	A05	21798836	3.84E-07*	9.15	A/G		○			√		●	●
Bn-A07-p3687800	A07	3687800	3.61E-07*-1.33E-07*	9.31–10.40	C/T		○			√		●	●
Bn-A08-p10625793	A08	10625793	1.16E-07*	10.65	A/G		○			√		●	
Bn-A08-p10996087	A08	10996087	2.74E-07*	9.63	C/T		○			√		●	
Bn-A08-p13693932	A08	13693932	1.35E-07*-6.06E-09**	11.32–11.52	A/G		○			√		●	●
Bn-C01-p17487034	C01	17487034	3.01E-07*	8.98	A/G		○			√		●	
Bn-C01-p17487264	C01	17487264	2.06E-07*	9.2	A/G		○			√		●	
Bn-C01-p17494228	C01	17494228	2.62E-08**	10.52	C/T		○			√		●	
Bn-C01-p17730780	C01	17730780	3.24E-07*	9.21	C/T		○			√		●	
Bn-C01-p17753079	C01	17753079	2.67E-07*	10.4	A/T		○			√		●	
Bn-C01-p17753141	C01	17753141	2.84E-07*	10.21	A/T		○			√		●	
Bn-C01-p17753145	C01	17753145	1.50E-07*	10.71	A/G		○			√		●	
Bn-C01-p17761899	C01	17761899	5.94E-08*	10.66	C/T		○			√		●	
Bn-C01-p17761901	C01	17761901	9.176E-08*	10.37	A/G		○			√		●	
Bn-C01-p17761933	C01	17761933	6.88E-08*	10.54	C/T		○			√		●	
Bn-C02-p33057504	C02	33057504	4.23E-07*-6.34E-08*	9.87–10.28	A/G		○			√		●	●
Bn-C03-p20819013	C03	2081903	1.26E-08**	8.92	G/T		○			√		●	
Bn-C03-p4949866	C03	4949866	1.94E-07*-4.85E-07*	11.13–12.55	A/T		○			√		●	●
Bn-C03-p38949187	C03	38949187	1.48E-07*	9.35	C/T		○			√		●	
Bn-C03-p38949208	C03	38949208	1.48E-07*	9.35	A/T		○			√		●	
Bn-C03-p38949217	C03	38949217	1.48E-07*	9.35	A/G		○			√		●	
Bn-C03-p38949237	C03	38949237	1.39E-07*	9.39	C/T		○			√		●	
Bn-C03-p38949409	C03	38949409	1.21E-07*	9.47	A/C		○			√		●	
Bn-C03-p38949491	C03	38949491	1.21E-07*	9.47	A/C		○			√		●	
Bn-C03-p40406280	C03	40406280	6.09E-08*	9.75	A/G		○			√		●	
Bn-C03-p40406328	C03	40406328	6.18E-08*	9.75	A/G		○			√		●	
Bn-C03-p40406358	C03	40406358	1.134E-07*	9.4	A/G		○			√		●	
Bn-C03-p40406652	C03	40406652	1.25E-07*	9.32	G/T		○			√		●	
Bn-C03-p46530629	C03	46530629	1.19E-07*	9.32	A/G		○			√		●	
Bn-C04-p723016	C04	723016	4.18E-07*	9.51	A/T		○			√		●	
Bn-C04-p32566552	C04	32566552	4.18E-08**	9.9	A/G		○			√		●	
Bn-C05-p3438659	C05	3438659	1.18E-07*-4.67E-07*	9.33–10.15	A/T		○			√		●	●
Bn-C05-p36767961	C05	36767961	4.52E-07*-1.83E-07*	9.30–10.24	A/C		○			√		●	●
Bn-C06-p10488893	C06	10488893	3.35E-07*-1.45E-07*	9.36–10.61	A/G		○			√		●	●
Bn-C08-p21679211	C08	21679211	2.334E-07*-1.44E-08**	11.26–12.88	C/T		○			√		●	●
Bn-C08-p21679214	C08	21679214	2.85E-07*-6.44E-09**	11.85–12.92	G/T		○			√		●	●
Bn-C08-p21679454	C08	21679454	2.94E-07*-1.85E-08**	11.18–12.60	A/G		○			√		●	●
Bn-C08-p21679515	C08	21679515	2.65E-07*-1.82E-08**	11.14–12.69	A/C		○			√		●	●
Bn-C08-p21679556	C08	21679556	2.79E-07*-1.94E-08**	11.15–12.65	C/T		○			√		●	●
Bn-A09-p122632	A09	122636	1.55E-07*	10.96	G/T	○				√			●
Bn-C09-p45893469	C09	45893469	4.12E-07*	10.83	A/G		○			√		●	
Bn-SA07-p363274	scaffoldA07_random	363274	2.88E-09**	11.83	C/T		○			√		●	

Continued

Table 3. continued

SNPs	Chromosome	Position	P value	R² (%)	Allele	Environment		Traits				Methods	
						JXAU	JXIRS	IFS	FFS	FP	MT	GLM	MLM
Bn-SA07-p363279	scaffoldA07_random	363279	3.06E-09**	11.8	A/G		○			√		•	
Bn-SA08-p1478242	scaffoldA08_random	1478242	8.21E-08*	10.11	A/C		○			√		•	
Bn-SC01-p1541365	scaffoldC01_random	1541365	8.85E-08*	10.64	A/G		○			√		•	
Bn-SC03-p643771	scaffoldC03_random	643771	7.81E-08*	9.58	A/C		○			√		•	
Bn-A10-p13390883	A10	13390883	4.53 E-07*	13.04	C/A	○					√	•	
Bn-C01-p21914037	C01	21914037	3.52 E-07*	13.23	A/G	○					√	•	
Bn-C03-p35944796	C03	35944796	4.29E-07*	7.27	A/G	○					√	•	
Bn-C04-p27101283	C04	27101283	2.94 E-07*	6.46	A/G	○					√	•	
Bn-C05-p29165950	rC05	29165950	1.75E-07*	6.78	A/G	○					√	•	
Bn-C07-p1361000	C07	1361000	4.37E-07*	7.57	C/T	○					√	•	
Bn-SA05-p1133338	scaffoldA05_random	1133338	3.73E-07*	7.34	C/T	○					√	•	
Bn-SA05-p1133563	scaffoldA05_random	1133563	3.25E-07*	7.41	A/G	○					√	•	
Bn-A07-p5248478	scaffoldA07_random	5248478	2.41E-07*	10.17	C/T	○					√	•	

*Significant SNP locus with $P<4.96E-07$.

**Highly significant SNP locus with $P<4.96E-08$. R^2 is the percentage of phenotypic variance explained by the SNP. ○ indicates the corresponding environment where the significant SNP locus located; √ indicates the corresponding trait that the significant SNP locus associated; ● indicates the corresponding model detecting the significantly associated-trait SNP locus.
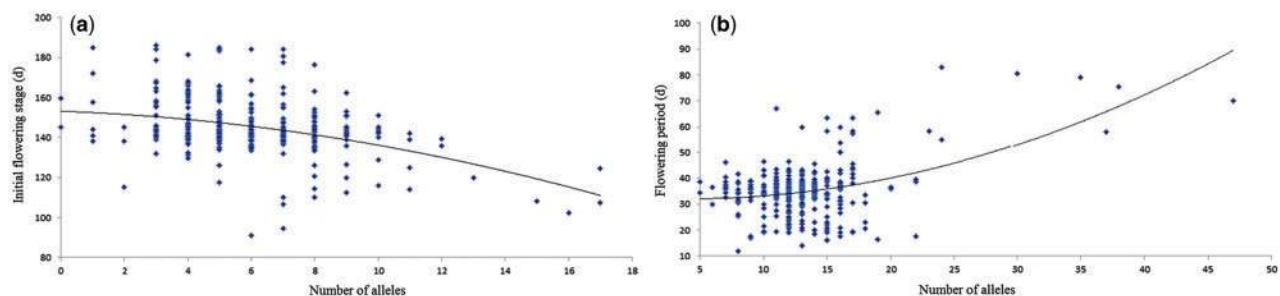

Figure 5. Analysis of numbers of highly favorable SNP alleles for early-flowering and long flowering period in *Brassica napus*.
Note: (a) The *X*-axis indicates the number of highly favorable SNP alleles for early-flowering and the *Y*-axis indicates the average IFS value in each accession; (b), the *X*-axis indicates the number of highly favorable SNP alleles for long-flowering and the *Y*-axis indicates average FP values in each accession.
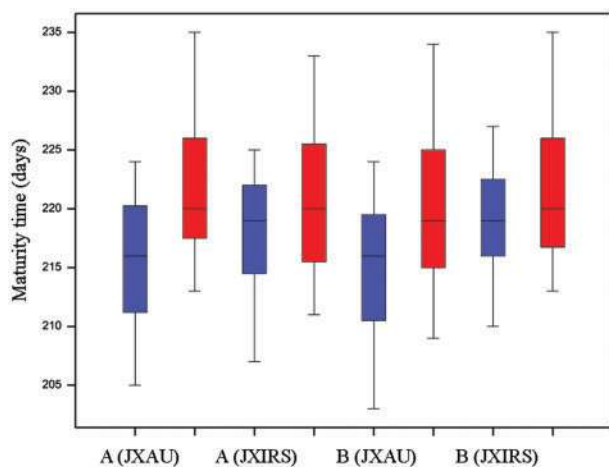

Figure 6. Boxplots showing maturity time for two genotypes carrying the C-allele (left) and the G-allele (right) for each location. A represents the SNP locus of *Bn-A02-23681432*, B represents the SNP locus of *Bn-A02-23875175*.

A total of 131 SNP loci associated with these traits were detected on 18 chromosomes (except A04), with a high average phenotypic variation for flowering time (9.36%) ranging from 3.27% to 13.17%, of which the greatest number of SNP loci was significantly associated

with FP. Many of these SNPs were also associated with IFS. From this, we can infer that SNP loci related to the initiation of flowering also mediate the flowering days. In comparison to bi-parental mapping population results using *in silico* mapping, eight SNPs detected in our study were consistent with previous flowering time and MT QTLs (in the range of 1 Mb) on chromosomes A02, A03, A05, A06, C03 and C08 [16,20,53,55] (Supplementary Table S18). In addition, based on the comparison of SNP regions, 37 SNP regions (within 200 kb) we detected were consistent with results of flowering time from the *Brassica* 60 K SNP array: 17 were reported by Li et al.,[56] 16 were reported by Roman et al.[46] and 5 were reported by Wang et al.[31] (Supplementary Table S19). Overall, at least 34 flowering QTLs in the current study were consistent with at least one QTL identified in one or more previous studies, and five SNP loci regions (*Bn-A02-6435246*, *Bn-A05-21493588*, *Bn-A06-7621426*, *Bn-SA03-2039163* and *Bn-SC03-643771*) were detected simultaneously by linkage mapping and association analysis. In addition, we found the SNP locus of *Bn-A03-25129078* is in the inner region of candidate gene of *GSBRNA2T00044315001* homologous to *CURLY LEAF* (*CLF*) of *A. thaliana*, while another five candidate genes for flowering time were within 10 kb of significant SNP loci (Supplementary Table S16). These findings strongly support the GWAS results and increase the credibility of the trait-associated SNP loci identified in our study. Furthermore, a total of 91 novel SNP loci from 16
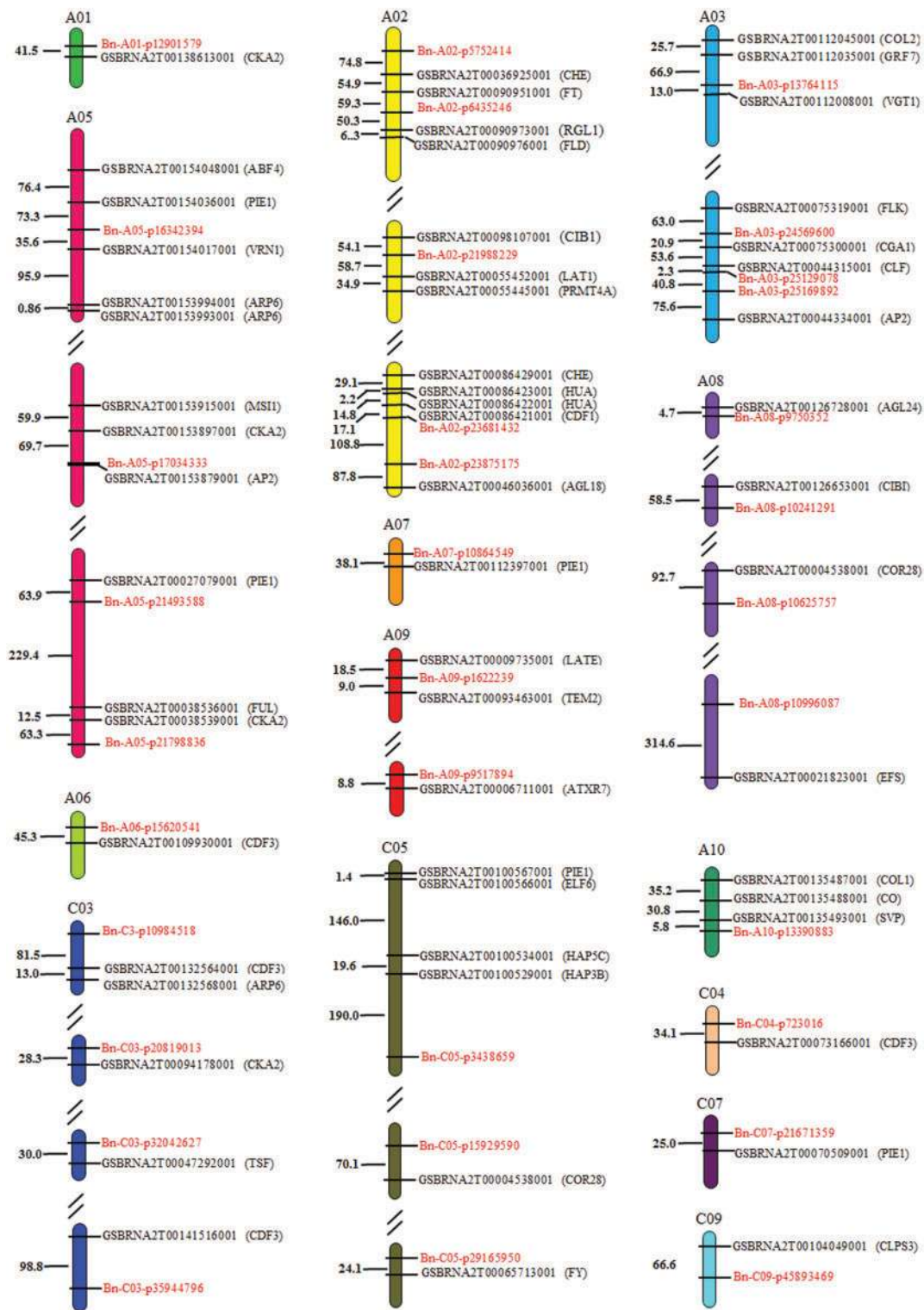
Figure 7. Distribution of candidate genes and their corresponding SNP loci associated with flowering time.

Note: The abbreviations of orthologous genes in *Arabidopsis thaliana* are shown in brackets after the candidate genes. Numbers represent the relative distances in the genome, 1 = 1 kb.

chromosomes were found in our study, 59 of which included 77 favourable alleles promoting early flowering and early maturity (Supplementary Tables S10 and S11), which might comprise new DNA markers for earliness breeding in rapeseed in the future.

Moreover, as the SNP markers used in this study were developed from sequencing, their position and alleles are known. Hence, based on our results, breeders can directly obtain valuable data and resources for further research in rapeseed.

It is well known that it is difficult to simultaneously improve the early maturity and yield of a crop by traditional breeding methods. Therefore, the excavation of favourable SNP alleles is necessary for improving the complicated earliness trait in rapeseed using molecular marker-assisted selection (MAS). Association mapping has played an important role in exploring the elite alleles of many agronomic traits in *B. napus* (seed yield, flowering time, seed oil content and fatty acid compositions) in recent years. [48,28,57] In the current study, the phenotypic effect value of each allele for four traits was evaluated to obtain 35 favourable alleles for early flowering and early maturing, and 90 for long FP. In fact, all favourable alleles for prolonging FP were consistent with alleles for promoting flowering, so we can infer that favourable alleles for prolonging FP may also produce positive effects in promoting flowering. Previously, pyramiding effects of favourable SNP alleles has proved useful in building disease resistance, increasing fruit yield and improving quality traits. [58–60] By analysing and comparing the favourable alleles for the four traits across the 300 accessions (Supplementary Tables S10–S13, Fig. 5), those which have more favourable alleles (such as 'Huayou 4' and 'Yuyou 2') might be considered as potential Germplasm resources for earliness breeding, and significant SNPs with favourable alleles can be used for MAS in rapeseed.

4.2 Mining of candidate genes to uncover the flowering time gene network and improve earliness in *B. napus*

Without doubt, the earliness of rapeseed largely depends on a complicated flowering network of genetic factors and their interaction with stimuli from the external environment. The genetic factors inducing the initiation of flowering are best elaborated in the model plant of *A. thaliana*,[10] where the regulation pathways for flowering time include intrinsic (autonomous, circadian clock, gibberellin) and extrinsic factors (vernalization, photoperiod and environmental temperature), and involve more than 100 flowering time genes. In this study, we identified 57 candidate genes of *B. napus* homologous to 39 flowering time genes of *A. thaliana* (e.g. *AGL24*, *FT*, *CO*, *SVP*, *FLD*, *FY*) in the vicinity of 33 significantly trait-associated SNP loci. These genes accounted for one-third of the known genetic and epigenetic regulators in the flowering time gene network, and 19 candidate genes homologous to 10 flowering time genes of *A. thaliana* (*TSF*, *CIB1*, *PIE1*, *VRN1*, *FUL*, *CKA2*, *ARP6*, *FY*, *ABF4* and *CDF3*) were detected near 14 novel SNP loci in our study. For these candidate genes for flowering time in *B. napus*, some of which play a positive role in promoting flowering, homologous genes of *A. thaliana* were as follows: *AGL24*, *CIB1*, *CLPS3*, *CO*, *RGL1*, *COR28*, *FLD*, *FLK*, *FT*, *FUL*, *FY*, *GRF7*, *HUA2*, *PRMT4A*, *TSF*, *VGT1*, *VRN1* and *AP2*; other candidate genes for delaying flowering were homologous to *EFS*, *LATE*, *SVP*, *AGL18*, *ABF4*, *ATXR7*, *CDF1*, *CDF3*, *CLF*, *CGA1*, *CKA2*, *ELF6*, *ARP6*, *PIE1* and *TEM2* in *A. thaliana* (Supplementary Table S17). Thus, it is reasonable to suppose that those genes homologous to flowering time genes of *A. thaliana* for promoting flowering may be considered as candidate genes for improving earliness via the regulation and control of early flowering time in *B. napus*.

Of the flowering time candidate genes detected in this study, gene *GSBRNA2T00135488001* adjacent to the SNP locus *Bn-A10-13390883* was homologous to *CONSTANS* (*CO*), temporal and spatial regulation of which is vital for photoperiod-dependent induction.[12] Four orthologues of the *A. thaliana CO* gene have previously been isolated on chromosomes A10 and C09 in *B. napus*.[61] The *GSBRNA2T00098107001* gene adjacent to *Bn-A02-21988229* and the

GSBRNA2T00126653001 gene adjacent to *Bn-A08-10241291* were homologous to *CRYPTOCHROME-INTERACTING BASIC-HELIX-LOOP-HELIX 1* (*CIB1*), which interacts with *CRYPTOCHROME 2* (*CRY2*) to promote *CRY2*-dependent floral initiation and also positively regulates *FLOWERING LOCUS T* (*FT*) expression in the photoperiod pathway in *A. thaliana*.[62] In addition, *EARLY FLOWERING 6* (*ELF6*) encodes a Jumonji N/C and zinc finger domain-containing protein that acts as a repressor in the photoperiod pathway, and its loss-of-function mutation causes early flowering[63]: we found its homologous candidate gene *GSBRNA2T00100566001* in the vicinity of *Bn-C05-3438659*. Many flowering time candidate genes are involved in vernalization pathways, as vernalization promotes flowering indirectly by histone modifications that submerge *FLOWERING LOCUS C* (*FLC*). At present, five regulator genes (*VIN3*, *VRN5*, *VRN1*, *VRN2* and *HPL1*) related to vernalization have been found in *A. thaliana*.[12] Candidate gene *GSBRNA2T00154017001* homologous to *VERNALIZATION 1* (*VRN1*) near SNP *Bn-A05-16342394*, which could repress *FLC* gene expression by regulating the chemical modification of histones,[64] and three additional candidate genes (*GSBRNA2T00153994001*, *GSBRNA2T00153993001* and *GSBRNA2T00132568001*) homologous to *ACTIN RELATED PROTEIN 6* (*ARP6*), which could act in the nucleus to modulate *FLC* gene expression by participating in chromatin histone 3 acetylation,[65,66] were found in the vicinity of SNPs *Bn-A05-16342394* and *Bn-C03-10984518* in this study. Although we did not find homologous for the vital regulator of flowering time *FLC* in the candidate regions in our study, *FLC* is not a unique target gene in the vernalization pathway, as *AGL19* and *AGL24* encoding MADX-box proteins have similar roles to *FLC*, where up-regulated expression can promote precocious flowering.[67,68] We did detect the gene *GSBRNA2T00126728001* homologous to *AGL24* [in the vicinity of SNP locus (*Bn-A08-9750352*)], which may play important roles in the downstream regulation of *SOC1* and upstream regulation of *LFY* in several floral pathways: this gene is firstly activated in shoot apical meristems at the stage of floral transition, after which expression is located in inflorescence and floral meristems. FLOWERING LOCUS D (*FLD*), FLOWERING LOCUS K (*FLK*) and *FY* are the most important flowering time genes in the autonomous flowering pathway, as these genes promote flowering indirectly by repressing the expression of *FLC*, but they have operate via different mechanisms. We found the candidate gene of *GSBRNA2T00090976001* homologous to *FLD* (a high acetylation transcriptional repressor of *FLC*[69]) 56.69 kb from the SNP locus *Bn-A02-6435246*. In addition, we also excavated the candidate genes *GSBRNA2T00075319001* (*Bn-A03-24569600*, 63.03 kb) and *GSBRNA2T00065713001* (*Bn-C05-29165950*, 24.13 kb) homologous to *FLK* and *FY*, respectively, both of which encode RNA-binding proteins known to affect flowering time by modulating the mRNA level of *FLC*.[70] In the GA (gibberellin) pathway, *RGA-LIKE 1* (*RGL1*) is known to be a repressor of the GA response pathway controlling flowering,[71] and we identified its homologous gene *GSBRNA2T00090973001* 50.39 kb from *Bn-A01-6435246*.

The floral integrator genes (*SOC1*, *FT* and *AGL24*) play important roles in activating floral meristem formation genes (such as *LFY*, *AP1*, *SEP3* and *FUL*). Besides AGL24 mentioned above, we also found the candidate gene *GSBRNA2T00090951001* homologous to *FLOWERING LOCUS T* (*FT*) closely linked with SNP marker *Bn-A02-p6435246*, which was also reported previously as a flowering time QTL.[16] Six orthologues of *Arabidopsis FT* have previously been identified in the genome of *B. napus*.[72] *FT* plays a vital role in the floral transition process as the floral integrator in the photoperiod pathway and as a signalling molecule from the leaves to the apical

meristem.[13,73] As well, gene *GSBRNA2T00038536001* on chromosome A05 was proved to be homologous to *FRUITFULL* (*FUL*) of *Arabidopsis*, a MADS-box transcription factor which may act as a molecular switch between the vegetative and reproductive states by forming *FUL-SVP* and *FUL-SOC1* heterodimers.[74] The *GSBRNA2T00044334001* gene on A03 and the *GSBRNA2T00153879001* gene on chromosome A05 were homologous to *APETALA 2* (*AP2*) of *Arabidopsis*, and play important roles in regulating flower development and specification of floral organ identity.[75,76] The *GSBRNA2T00055445001* gene on chromosome A02 was homologous to *PROTEIN ARGININE METHYL TRANSFERASE 4A* (*PRMT4A*) of *Arabidopsis*, which is known to directly induce the expression of floral repressor *AGAMOUS-LIKE15* (*AGL15*) and to repress the transcription of floral activators such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*).[77]

Nevertheless, many important flowering time genes in the genetic network were not found in the current study, such as *FLC*, *SOC1*, *FRI*, *FD*, *SOC1* and *LFY*, among others. We think the main reason for this is that most of the accessions (257 of 300) in the association population are semi-winter varieties. Therefore, significantly trait-associated SNP loci were mainly from semi-winter types, and hence candidate genes for vernalization sensitivity would not be easy to find. For example, the central flowering time suppressor *FLC* in the vernalization pathway was not found in the candidate regions. In addition, the similar thermo-light conditions of the two environments make it difficult to explore flowering time candidate genes related to photoperiod and ambient temperature pathways. It is also possible that some important loci associated with flowering time genes were omitted due to failure to satisfy the high *P* value threshold (<4.96E10-7) used to identify true marker–trait associations.

In this study, we investigated the phenotypes of four traits related to earliness of *B. napus* in two environments, based on 201,187 SNP markers developed from SLAF-seq. We performed a genome-wide association analysis of four traits across 300 rapeseed inbred lines, and 131 SNPs significantly associated with these traits were detected on 18 chromosomes using GLM and MLM analyses. Highly favourable alleles for promoting flowering time and prolonging the FP were excavated. Moreover, we identified 57 flowering time candidate genes in the vicinity of 33 SNP loci significantly associated with these traits. In summary, we present a series of exploratory analyses of earliness loci and flowering time candidate genes based on a GWAS. This GWAS approach showed great power in uncovering genetic variation in flowering time in *B. napus*, enhancing our knowledge of the molecular mechanisms controlling flowering in rapeseed. The elite alleles identified that contribute to earliness in *B. napus* can be directly applied to the targeted breeding of earliness in rapeseed, facilitating commercial rapeseed cultivation across greater regions worldwide.

Conflict of interest

None declared.

Funding

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Nagaharu, U. 1935, Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization, *Jpn. J. Bot.*, **7**, 389–452.
2. Ziolkowski, P.A., Kaczmarek, M., Babula, D. and Sadowski, J. 2006, Genome evolution in *Arabidopsis/Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints, *Plant J.*, **47**, 63–74.
3. Friedt, W. and Snowdon, R. 2009, Oilseed rape. *Oil Crops*, Springer, pp. 91–126.
4. Saeidnia, S. and Gohari, A. R. 2012, Importance of *Brassica napus* as a medicinal food plant, *J. Med. Plants Res.*, **6**, 2700–3.
5. Liu, H. 2000, Genetics and breeding in rapeseed. *Chinese Agricultural Universitatis*, Beijing, 144–77.
6. Fu, T. 2000, Breeding and utilization of rapeseed hybrid. *Hubei Science Technology*, Hubei, 167–9.
7. Prakash, S., Wu, X.M. and Bhat, S. 2011, History, evolution, and domestication of *Brassica* crops, *Plant Breed. Rev.*, **35**, 19–84.
8. Rahman, H., Bennett, R.A. and Kebede, B. 2017, Mapping of days to flower and seed yield in spring oilseed *Brassica napus* carrying genome content introgressed from *Brassica oleracea*, *Mol. Breed.*, **37**, 5.
9. Fu, D., Jiang, L. and Mason, A.S. 2016, Research progress and strategies for multifunctional rapeseed: a case study of China, *Integr. Agr.*, **15**, 1673–84.
10. Blümel, M., Dally, N. and Jung, C. 2015, Flowering time regulation in crops—what did we learn from *Arabidopsis*? *Curr. Opin. Biotechnol.*, **32**, 121–9.
11. Jung, C. and Müller, A.E. 2009, Flowering time control and applications in plant breeding, *Trends Plant Sci.*, **14**, 563–73.
12. Srikanth, A. and Schmid, M. 2011, Regulation of flowering time: all roads lead to Rome, *Cell. Mol. Life Sci.*, **68**, 2013–37.
13. Wigge, P.A. 2013, Ambient temperature signalling in plants, *Curr. Opin. Plant Biol.*, **16**, 661–6.
14. Ferreira, M., Satagopan, J., Yandell, B., Williams, P. and Osborn, T. 1995, Mapping loci controlling vernalization requirement and flowering time in *Brassica napus*, *Theoret. Appl. Genet.*, **90**, 727–32.
15. Zhao, J., Becker, H., Ding, H., Zhang, Y., Zhang, D. and Ecke, W. 2005, QTL of three agronomically important traits and their interactions with environment in a European × Chinese rapeseed population, *Yi Chuan Xue Bao*, **32**, 969–78.
16. Udall, J. A., Quijada, P. A., Lambert, B. and Osborn, T. C. 2006, Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm, *Theor. Appl. Genet.*, **113**, 597–609.
17. Long, Y., Shi, J. and Qiu, D. 2007, Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with Arabidopsis, *Genetics*, **177**, 2433–44.
18. Mei, D., Wang, H., Hu, Q., Li, Y., Xu, Y. and Li, Y. 2009, QTL analysis on plant height and flowering time in *Brassica napus*, *Plant Breed.*, **128**, 458–65.
19. Wang, N., Qian, W., Suppanz, I., et al. 2011, Flowering time variation in oilseed rape (*Brassica napus* L.) is associated with allelic variation in the *FRIGIDA* homologue BnaA, FRI. a, *J. Exp. Bot.*, **62**, 5641–58.
20. Shi, J., Li, R., Zou, J., Long, Y., Meng, J. and Hansson, B. 2011, A dynamic and complex network regulates the heterosis of yield-correlated traits in rapeseed (*Brassica napus* L.), *PLoS One*, **6**, e21645.
21. Würschum, T., Liu, W., Maurer, H.P., Abel, S. and Reif, J.C. 2012, Dissecting the genetic architecture of agronomic traits in multiple segregating populations in rapeseed (*Brassica napus* L.), *Theor. Appl. Genet.*, **124**, 153–61.
22. Raman, H., Raman, R., Eckermann, P., et al. 2013, Genetic and physical mapping of flowering time loci in canola (*Brassica napus* L.), *Theor. Appl. Genet.*, **126**, 119–32.

23. Flint-Garcia, S. A., Thornsberry, J. M. and Buckler IV, E. S. 2003, Structure of linkage disequilibrium in plants, *Annu. Rev. Plant Biol.*, **54**, 357–74.

24. Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., et al. 2010, Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines, *Nature*, **465**, 627–31.

25. Li, H., Peng, Z., Yang, X., et al. 2013, Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels, *Nat. Genet.*, **45**, 43–50.

26. Huang, X., Zhao, Y., Li, C., et al. 2011, Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm, *Nat. Genet.*, **44**, 32–9.

27. Li, F., Chen, B., Xu, K., et al. 2014, Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.), *DNA Res.*, **21**, 355–67.

28. Liu, S., Fan, C., Li, J., et al. 2016, A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*, *Theor. Appl. Genet.*, **129**, 1203–15.

29. Xu, L., Hu, K., Zhang, Z., et al. 2016, Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.), *DNA Res.*, **23**, 43–52.

30. Schiessl, S., Iniguez-Luy, F., Qian, W. and Snowdon, R.J. 2015, Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*, *BMC Genomics*, **16**, 737.

31. Wang, N., Chen, B., Xu, K., et al. 2016, Association mapping of flowering time QTLs and insight into their contributions to rapeseed growth habits, *Front. Plant Sci.*, **7**, 338–48.

32. Ganal, M.W., Wieseke, R., Luerssen, H., et al. 2014, High-throughput SNP profiling of genetic resources in crop plants using genotyping arrays. *Genomics Plant Genetic Resources*, Springer, pp. 113–30.

33. Zhou, Q., Zhou, C., Zheng, W., et al. 2017, Genome-wide SNP markers based on SLAF-seq uncover breeding traces in rapeseed (*Brassica napus* L.), *Front. Plant Sci.*, **8**, 648–59.

34. Kong, F. 2005, *Quantitative Genetics in Plants*. Beijing, China.

35. Tang, Q.Y. and Zhang, C.X. 2013, Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research, *Insect Sci.*, **20**, 254–60.

36. Shi, J., Zhan, J., Yang, Y., et al. 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**, 10–1038.

37. Murray, M. and Thompson, W. F. 1980, Rapid isolation of high molecular weight plant DNA, *Nucleic Acids Res.*, **8**, 4321–6.

38. Sun, X., Liu, D., Zhang, X., et al. 2013, SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing, *PloS One.*, **8**, e58700.

39. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. 2007, TASSEL: software for association mapping of complex traits in diverse samples, *Bioinformatics*, **23**, 2633–5.

40. Alexander, D.H., Novembre, J. and Lange, K. 2009, Fast model-based estimation of ancestry in unrelated individuals, *Genome Res.*, **19**, 1655–64.

41. Hardy, O.J. and Vekemans, X. 2002, SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels, *Mol. Ecol. Notes*, **2**, 618–20.

42. Ginestet, C. 2011, Ggplot2: elegant graphics for data analysis, *J. R. Stat. Soc. A*, **174**, 245–6.

43. Turner, S.D. 2014, Qqman: an R package for visualizing GWAS results using QQ and manhattan plots, *BioRxiv*, 005165.

44. Benjamini, Y. and Hochberg, Y. 1995, Controlling the false discovery rate: a practical and powerful approach to muRAMANltiple testing, *J. R. Stat. Soc. B.*, 289–300.

45. Lü, H.Y., Liu, X.F., Wei, S.P. and Zhang, Y.M. 2011, Epistatic association mapping in homozygous crop cultivars, *PLoS One.*, **6**, e17773.

46. Raman, H., Raman, R., Coombes, N., et al. 2015, Genome-wide association analyses reveal complex genetic architecture underlying natural variation for flowering time in canola, *Plant Cell Environ*, **39**, 1228–39.

47. Barrett, J.C., Fry, B., Maller, J., et al. 2005, Haploview: analysis and visualization of LD and haplotype maps, *Bioinformatics*, **21**, 263–5.

48. Cai, D., Xiao, Y., Yang, W., et al. 2014, Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.), *Theor. Appl. Genet.*, **127**, 85–96.

49. Chalhoub, B., Denoeud, F., Liu, S., et al. 2014, Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome, *Science*, **345**, 950–3.

50. Zhou, Q., Fu, D., Mason, A.S., Zeng, Y., Zhao, C. and Huang, Y. 2014, In silico integration of quantitative trait loci for seed yield and yield-related traits in *Brassica napus*, *Mol. Breed.*, **33**, 881–94.

51. Lee, J. M. and Sonnhammer, E. L. 2003, Genomic gene clustering analysis of pathways in eukaryotes, *Genome Res.*, **13**, 875–82.

52. Cruz, V.M.V., Luhman, R., Marek, L.F., et al. 2007, Characterization of flowering time and SSR marker analysis of spring and winter type *Brassica napus* L. germplasm, *Euphytica*, **153**, 43–57.

53. Mahmood, T., Rahman, M.H., Stringam, G.R., Yeh, F. and Good, A.G. 2007, Quantitative trait loci for early maturity and their potential in breeding for earliness in *Brassica juncea*, *Euphytica*, **154**, 101–11.

54. Shi, J., Li, R., Qiu, D., et al. 2009, Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*, *Genetics*, **182**, 851–61.

55. Quijada, P.A., Udall, J.A., Lambert, B. and Osborn, T.C. 2006, Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm, *Theor. Appl. Genet.*, **113**, 549–61.

56. Li, L., Long, Y., Zhang, L., et al. 2015, Genome wide analysis of flowering time trait in multiple environments via high-throughput genotyping technique in *Brassica napus* L, *PLoS One.*, **10**, e0119425.

57. Gacek, K., Bayer, P.E., Bartkowiak-Broda, I., et al. 2017, Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*, *Front. Plant Sci.*, **7**, 2062.

58. Werner, K., Friedt, W. and Ordon, F. 2005, Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (*BaMMV, BaYMV, BaYMV-2*), *Mol. Breed.*, **16**, 45–55.

59. Sacco, A., Di, M.A., Lombardi, N., et al. 2013, Quantitative trait loci pyramiding for fruit quality traits in tomato, *Mol. Breed.*, **31**, 217–22.

60. Zhang, B., Li, W., Chang, X., et al. 2014, Effects of favorable alleles for watersoluble carbohydrates at grain filling on grain weight under drought and heat stresses in wheat, *PLoS One.*, **9**, e102917.

61. Robert, L.S., Robson, F., Sharpe, A., et al. 1998, Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*, *Plant Mol. Biol.*, **37**, 763–72.

62. Liu, H., Wang, Q., Liu, Y., et al. 2013, *Arabidopsis* CRY2 and ZTL mediate blue-light regulation of the transcription factor CIB1 by distinct mechanisms, *Proc. Natl. Acad. Sci. USA.*, **110**, 17582–7. P.

63. Noh, B., Lee, S.H., Kim, H.J., et al. 2004, Divergent roles of a pair of homologous Jumonji/Zinc-Finger–class transcription factor proteins in the regulation of *Arabidopsis* flowering time, *Plant Cell*, **16**, 2601–13.

64. Levy, Y.Y., Mesnage, S., Mylne, J.S., et al. 2002, Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control, *Science*, **297**, 243–6.

65. March Díaz, R., García Domínguez, M., Lozano-Juste, J., et al. 2007, Histone *H2A. Z* and homologues of components of the *SWR1* complex are required to control immunity in *Arabidopsis*, *Plant J.*, **53**, 475–87.

66. Kumar, S.V. and Wigge, P.A. 2010, *H2A.* Z-containing nucleosomes mediate the thermosensory response in *Arabidopsis*, *Cell*, **140**, 136–47.

67. Yu, H., Xu, Y., Tan, E.L., et al. 2002, *AGAMOUS-LIKE 24*, a dosage-dependent mediator of the flowering signals, *Proc. Natl. Acad. Sci. USA.*, **99**, 16336–41.

68. Schönrock, N., Bouveret, R., Leroy, O., et al. 2006, Polycomb-group proteins repress the floral activator *AGL19* in the *FLC*-independent vernalization pathway, *Gene Dev.*, **20**, 1667–78.

69. He, Y.H. 2009, Control of the transition to flowering by chromatin modifications, *Mol. Plant.*, **2**, 554–64.

70. Quesada, V., Dean, C. and Simpson, G.G. 2005, Regulated RNA processing in the control of *Arabidopsis* flowering, *Int. J. Dev. Biol.*, **49**, 773–80.

71. Galvão, V.C., Horrer, D., Küttner, F., et al. 2012, Spatial control of flowering by *DELLA* proteins in *Arabidopsis thaliana*, *Development*, **139**, 4072–82.

72. Wang, J., Long, Y., Wu, B., et al. 2009, The evolution of *Brassica napus FLOWERING LOCUST* paralogues in the context of inverted chromosomal duplication blocks, *BMC Evol. Biol.*, **9**, 271.

73. Corbesier, L., Vincent, C., Jang, S., et al. 2007, FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*, *Science*, **316**, 1030–3.

74. Balanzà, V., Martínez-Fernández, I. and Ferrándiz, C. 2014, Sequential action of *FRUITFULL* as a modulator of the activity of the floral regulators *SVP* and *SOC1*, *J. Exp. Bot.*, ert482.

75. Chen, X. 2004, A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development, *Science*, **303**, 2022–5.

76. Yant, L., Mathieu, J., Dinh, T.T., et al. 2010, Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor *APETALA2*, *Plant Cell*, **22**, 2156–70.

77. Niu, L., Zhang, Y., Pei, Y., Liu, C. and Cao, X. 2008, Redundant requirement for a pair of *PROTEIN ARGININE METHYLTRANSFERASE4* homologs for the proper regulation of Arabidopsis flowering time, *Plant Physiol.*, **148**, 490–503.