ORIGINAL ARTICLE

Non-vernalization requirement in Chinese kale caused by loss of *BoFLC* and low expressions of its paralogs

Qiwei Tang[1] · Hanhui Kuang[1] · Changchun Yu[1] · Guanghui An[1] · Rong Tao[1] · Weiyi Zhang[1] · Yue Jia[1]

Abstract

Key message **We identified the loss of *BoFLC* gene as the cause of non-vernalization requirement in *B. oleracea*. Our developed codominant marker of *BoFLC* gene can be used for breeding program of *B. oleracea* crops.**

Abstract Many species of the Brassicaceae family, including some Brassica crops, require vernalization to avoid pre-winter flowering. Vernalization is an unfavorable trait for Chinese kale (*Brassica oleracea* var. *chinensis Lei*), a stem vegetable, and therefore it has been lost during its domestication/breeding process. To reveal the genetics of vernalization variation, we constructed an F_2 population through crossing a Chinese kale (a non-vernalization crop) with a kale (a vernalization crop). Using bulked segregant analysis (BSA) and RNA-seq, we identified one major quantitative trait locus (QTL) controlling vernalization and fine-mapped it to a region spanning 80 kb. Synteny analysis and PCR-based sequencing results revealed that compared to that of the kale parent, the candidate region of the Chinese kale parent lost a 9,325-bp fragment containing *FLC* homolog (*BoFLC*). In addition to the *BoFLC* gene, there are four other *FLC* homologs in the genome of *B. oleracea*, including *Bo3g005470*, *Bo3g024250*, *Bo9g173370,* and *Bo9g173400*. The qPCR analysis showed that the *BoFLC* had the highest expression among the five members of the *FLC* family. Considering the low expression levels of the four paralogs of *BoFLC*, we speculate that its paralogs cannot compensate the function of the lost *BoFLC,* therefore the presence/absence (PA) polymorphism of *BoFLC* determines the vernalization variation. Based on the PA polymorphism of *BoFLC*, we designed a codominant marker for the vernalization trait, which can be used for breeding programs of *B. oleracea* crops.

Introduction

Chinese kale (*Brassica oleracea var. chinensis Lei*) is a Brassica vegetable widely planted in Southern China and Southeast Asia (Lei et al. 2017; Qian et al. 2016). The stem and leaves of Chinese kale are popularly consumed owing to its flavors, vivid colors, rich anticarcinogenic and antioxidative components such as vitamin C, total phenolics, carotenoids, and glucosinolates (Sun et al. 2011, 2012; Wei et al. 2011; Wu et al. 2017). The bolting stems of Chinese kale are generally harvested and consumed, and consequently, Chinese kale has been bred for early bolting without vernalization.

For plants, the timing of reproduction is mainly governed by flowering time regulation. In temperate climates, two most important cues to induce flowering are a period of extended cold (also known as the vernalization pathway) and day length (the photoperiod pathway) (Blumel et al. 2015). The vernalization pathway is a complex regulatory system associated with transcriptional and epigenetic regulations (Song et al. 2013). In Arabidopsis, the vernalization pathway promotes flowering in response to extended exposure to low temperatures (Searle et al. 2006). The natural variation of flowering time in Arabidopsis can be explained by two loci, *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) (Burn et al. 1993; Clarke and Dean 1994; Johanson et al. 2000; Koornneer et al. 1994; Michaels and Amasino 1999; Sheldon et al. 1999). *FLC* encoding a MADS-box transcription factor inhibits expression of the central flowering regulator *FLOWERING LOCUS T* (*FT*) (Sheldon et al. 2000). FLC also binds to the promoters and represses the expressions of several other important flowering genes, such as *FLOWERING LOCUS D* (*FD*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), and *TEMPRANILLO 1*

✉ Yue Jia
   jiayue@webmail.hzau.edu.cn

1 Key Laboratory of Horticultural Plant Biology, Ministry
   of Education, Key Laboratory of Horticultural Crop Biology,
   College of Horticulture and Forestry Sciences, Huazhong
   Agricultural University, Wuhan 430070, China

(*TEM1*) (Deng et al. 2011; Searle et al. 2006). *FRI* and *VERNALIZATION INSENSITIVE 3* (*VIN3*) are two critical genes regulating the expression of *FLC* (Johanson et al. 2000; Michaels and Amasino 1999; Sheldon et al. 1999). *FRI* is responsible for a high production of *FLC* and *VIN3* genes at both protein and transcriptional levels during vernalization. The vernalization process overrides the FRI-mediated control of *FLC*, resulting in the repression of transcriptional activity of *FRI* and the promotion of flowering initiation (Mempel et al. 2004; Sheldon et al. 2000).

The main vernalization components are conserved in Brassicaceae species, and their paralogs also have similar functions in regulating flowering time. Variation in paralogs of *FLC* has been reported to be related to flowering time in *B. rapa*, *B. napus*, and *B. oleracea* (Hou et al. 2012; Irwin et al. 2016; Okazaki et al. 2007; Osborn et al. 1997; Pires et al. 2004; Razi et al. 2008; Ridge et al. 2015; Schiessl et al. 2017; Tadege et al. 2001; Wu et al. 2012; Yuan et al. 2009; Zhao et al. 2010). There are four homologs of *FLC* in *B. rapa*, including *BrFLC1*, *BrFLC2*, *BrFLC3*, *BrFLC5*. A naturally occurring InDel variation in *BrFLC2* and a splicing site mutation in *BrFLC1* have also been found to contribute to flowering time variation in *Brassica rapa* (Wu et al. 2012; Yuan et al. 2009). A study based on genetic mapping and QTL analysis has demonstrated that *BrFLC1*, *BrFLC2,* and *BrFLC5* play important roles in vernalization, suggesting that these *FLCs* control flowering response in a dosage-dependent manner (Schranz et al. 2002b). Rapeseed (*Brassica napus*) with a highly duplicated genome carries nine copies of *Bna.FLC,* and insertion of a Tourist-like MITE into the upstream region of the *BnFLC.A10* gene affects vernalization requirement (Hou et al. 2012). Four *FLC* homologs (*BoFLC1*, *BoFLC3*, *BoFLC4*, *BoFLC5*) have been identified from *B. oleracea* (Lin et al. 2005; Schranz et al. 2002b). *BoFLC.C2* co-located with a flowering time QTL has been reported to be responsible for the late flowering trait in broccoli (*Brassica oleracea var. italica*) (Okazaki et al. 2007; Zhao et al. 2010).

Non-vernalization is an important agronomic trait for Chinese Kale, a vegetable with bolting stem harvested and consumed. In this study, we explored the genetic mechanism underlying vernalization variation in Chinese Kale and kale based on BSA + RNA-seq. The candidate gene was identified by synteny analysis of the candidate region and further verified through the comprehensive analysis of its paralogs. This study provides insight into the genetic and molecular mechanisms of vernalization in *B. oleracea*. The identified genetic variations can be used in marker-assisted genetic breeding of *B. oleracea* crops.

Materials and methods

Materials used in this study

The *B. oleracea* var. *Alboglabra* and *B. oleracea* var. *acephala* were used as parents to generate segregating population in this work. Cultivars, parental lines, F_1 hybrid, F_2, and advanced generations were grown in the field on the campus of Huazhong Agriculture University, Wuhan, China. The seeds of *B. oleracea* cultivars were acquired from National Center for Vegetable Improvement (Central China) or bought from commercial websites of JingDong (https://www.jd.com/).

Identification of vernalization variation and data analysis

F_1 plants were self-pollinated to produce F_2 population. A total of 107 individuals in F_2 population were planted in September, 2016 in Wuhan, China. Flowering status was investigated on December 27, 2016. The plants flowering before December 27, 2016 were defined as non-vernalization plants, and the plants flowering after December 27, 2016 were designated as vernalization plants. Three F_3 families were planted in September, 2017 in Wuhan, China. Flowering status was investigated in January 2018. One F_4 family (generated from one of the three F_3 families) with 268 individuals was planted in March, 2018. Flowering status was investigated in May, 2018.

ANOVA and phenotypic variation explained (PVE) analysis were conducted using R package 'car' (version 3.0–11) (https://CRAN.R-project.org/package=car).

Bulked segregant analysis (BSA) and RNA-seq

BSA in combination with RNA-seq (BSR) was used to identify QTL controlling flowering time following the procedures reported previously (Yu et al. 2020). In this study, the leaves from 20 four-month-old early flowering plants in an F_2 segregating population were mixed as an early flowering pool, and the leaves from 20 late flowering plants were mixed as late flowering pool. Total RNA was extracted from the two pools using TransZol reagent (TRANSGEN, Beijing, China). RNA-seq was performed on Illumina Hiseq2500 platform, and approximately 6 Gb clean data were obtained from each pool. Clean data were aligned to the *B. oleracea* genome TO1000 V2.1 (Liu et al. 2014) using Bowtie software (version 1.3.0) (Langmead et al. 2009). SNP calling was performed using SAMtools (version 1.7) (Li et al. 2009). Low-quality SNPs with mapping quality value < 30, read depth < 10, or base quality value < 20 were excluded.

SNP-index was used to identify the target region for early/late flowering trait. SNP-index was calculated by subtracting the allelic SNP frequency of the early flowering pool from that of the late flowering pool. Average SNP-index was calculated using a 900 kb sliding window with a step size of 400 kb, and was plotted along the nine chromosomes of *B. oleracea*. Finally, under the null hypothesis of no QTL, the 95% statistical confidence intervals of SNP-index were calculated as described previously (Takagi et al. 2013). In the plot, a peak may represent a region harboring a gene related to flowering time. The cleaved amplified polymorphic sequence (CAPS) markers were designed in the candidate region, and were used to screen the population and obtain causal genes by fine mapping.

PCR amplification of a large fragment

The genomic DNA of the parent plants was extracted from leaves using cetyltrimethylammonium bromide (CTAB) method (Porebski et al. 1997). Each 50 µL PCR mixture contained 1 µL of DNA template (100 ng/µL), 1 µL of each primer (10 µM), 22 µL of ddH$_2$O, and 25 µL PrimeSTAR GXL Premix (Takara Bio, Japan). The PCR program included 32 cycles of 98 °C for 10 s, 57 °C for 15 s, and 68 °C for 10 min. The PCR products were analyzed using 1% agarose gel electrophoresis. Purified PCR fragments were sequenced by commercial company Tsingke Biotechnology (Wuhan, China) through primer-walking method.

Sequence analysis

We analyzed the syntenic regions of the genome in *B. oleracea* species (the 'TO1000' reference genome, version 2.1) (Liu et al. 2014) and other species in Brassicaceae, including *B. rapa* (the 'Chiifu-401-42' reference genome, version 1.5) (Wang et al. 2011), *B. napus* (the Darmor-bzh reference genome, version 4.1) (Chalhoub et al. 2014), and Arabidopsis (version TAIR11) (Garcia-Hernandez et al. 2002) using BLAST (version 2.11.0) (Buchfink et al. 2015) with parameter "blastp -e 10–10". Primers used in this study were designed using PRIMER3 (version v.0.4.0) (Untergasser et al. 2012) (supplementary dataset 3). GENEIOUS (version 7.0.9; Biomatters Ltd, Auckland, New Zealand) was used to analyze sequence divergence among *FLC* homologs in *B. oleracea*.

Construction of phylogenetic tree

We obtained *FLC* gene identities (IDs) from previous studies of Brassicaceae (Calderwood et al. 2021; Lin et al. 2005; Schranz et al. 2002a). The amino acid sequences of the *FLC* homologs in different species were retrieved from EnsemblPlants database (http://plants.ensembl.org/index.

html) using *FLC* gene IDs. The protein sequences were aligned using Clustal X (version 2.1) (Larkin et al. 2007). A maximum-likelihood phylogenetic tree was constructed using MEGA 7.0 (Kumar et al. 2016). Bootstrapping was performed with 1000 replications.

Gene expression analysis

Total RNA was extracted from leaves using TransZol reagent (TRANSGEN, China, Beijing) following the manufacturer's instructions. Total RNA was treated with DNase I (Thermo, Waltham, USA) to remove the contaminated genomic DNA. The cDNA was synthesized using TransScript cDNA Synthesis SuperMix (TRANSGEN, Beijing, China). All reactions were performed using ChamQ SYBR qPCR Master Mix (Vazyme biotech, Nanjing, China) with a reaction system containing 10.0 µL of 2×ChamQ SYBR qPCR Master Mix, 0.4 µL of primers, 1.0 µL of cDNA, and 8.2 µL of ddH$_2$O. Quantitative real-time PCR (qRT-PCR) was conducted on QuantStudioTM 6 Real-Time PCR System (Thermo). The reactions were performed with three biological replicates and three technical replicates. The transcript levels of genes in different individuals were analyzed using the $2^{-\Delta\Delta CT}$ method and normalized with *actin* as an internal control. (Livak and Schmittgen 2001). The primers used for qRT-PCR assays were listed in the supplemental dataset 3.

Results

Genetic analysis of flowering time using BSA in combination with RNA-seq (BSR)

As a vegetable with bolting stem harvested and consumed, Chinese kale has been bred for early bolting. In contrast, kale, either as a vegetable or ornamental crop, requires vernalization. To genetically dissect flowering time variations between the two closely related crops, a Chinese kale cultivar was crossed with a kale cultivar (Fig. 1A), and the F$_1$ hybrids were selfed to generate an F$_2$ segregating population. BSR method was used to identify QTL controlling flowering time, and several peaks were observed in the plot. Each peak was considered to represent a QTL controlling flowering variation in F$_2$ population (Fig. 1C).

We named the QTL on chromosome 2, 5, and 9 as *Kale Vernalization Gene 1–3* (*KVG1-3*), respectively. ANOVA indicated no interactions among the three loci ($p = 0.47$, 0.86, 0.88), and *KVG1, KVG2* and *KVG3* explained 18.44%, 6.30%, and 2.08% of vernalization phenotypic variation, respectively. We concluded that in the F$_2$ segregating population, the vernalization was a quantitative trait controlled by at least three loci, and that *KVG1* was a major QTL for vernalization (Fig. 1C).
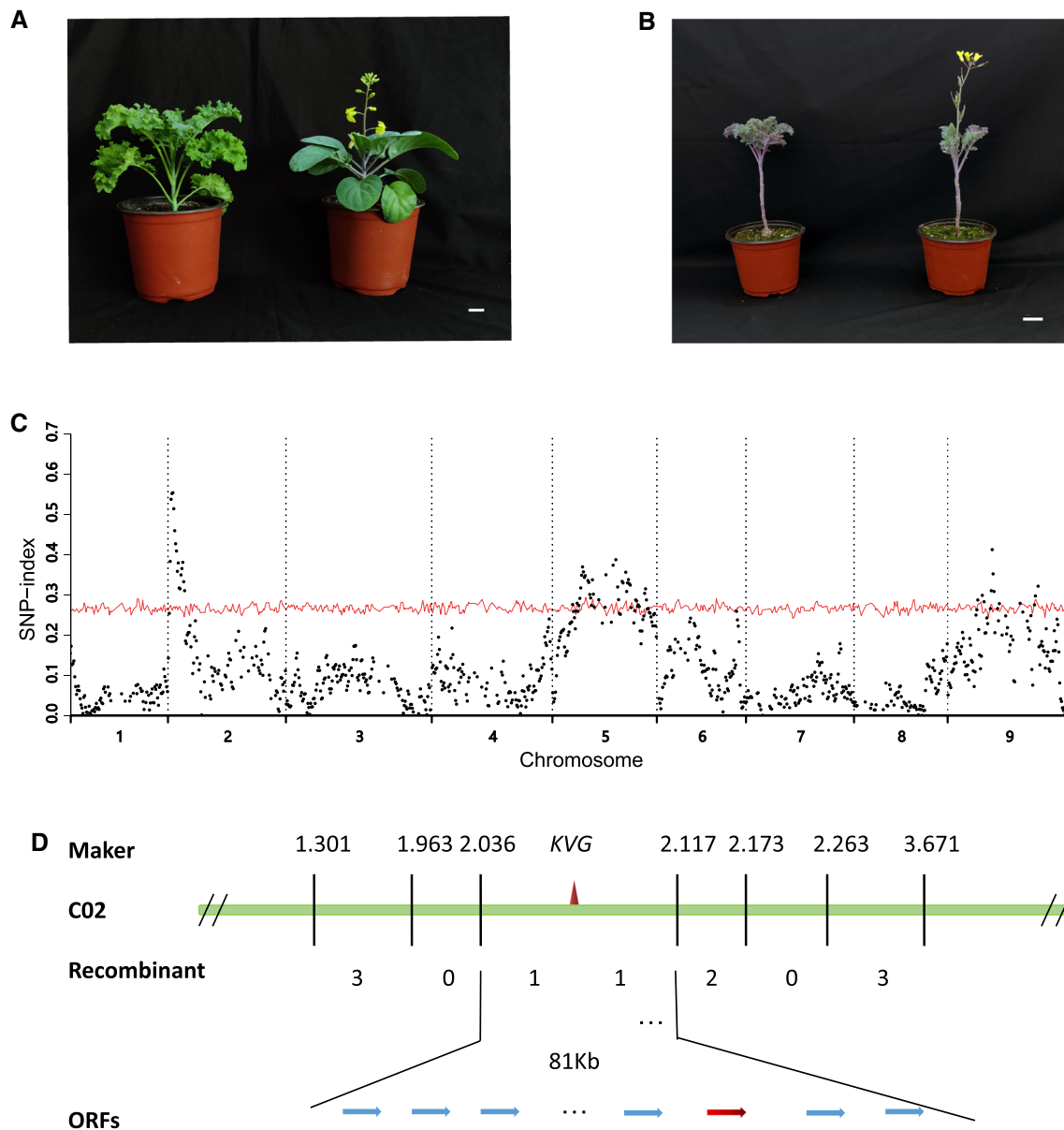
Fig. 1 Genetic mapping of genes affecting flowering time. **A** Parents used for constructing the F_2 population. Kale, left panel; Chinese kale, right panel. Bar = 2 cm. **B** Flowering individual and non-flowering individual in an F_4 family. Bar = 2 cm. **C** Plot of SNP-index along 9 chromosomes of cabbage. The *x*-axis represents the nine chromosomes and the y-axis represents the average SNP-index in each sliding window. The red curves represent statistical confidence intervals under the null hypothesis of no QTL at $p < 0.05$. **D** Fine mapping of *KVG1*. The numbers below the horizontal line and between two markers refer to the number of recombinants from 587 individuals

Construction of single-gene segregating population for vernalization trait

To clone the major QTL *KVG1* located on chromosome 2, we selected several F_2 individuals which are heterozygous at *KVG1* and homozygous at other QTL. These individuals were selfed to generate F_3 families in which only *KVG1* controls flowering time variation. The phenotype analysis of three F_3 families showed that flowering time was a qualitative trait in F_3 families, either flowering or non-flowering

before the winter, In January 2018, 78 individuals flowered and 231 individuals did not flower with a ratio of 1:3 ($\chi^2 = 0$, $p = 1$). The late flowering plants did not blossom until March 2018, showing typical vernalization characteristic. To further test whether the flowering trait was consistent with vernalization, we planted one F_4 family (generated from one of the three F_3 families) with 268 individuals in March, 2018. Of them, 73 individuals flowered in May, 2018, whereas 195 individuals did not flower until March 2019 with a ratio of 1:3 ($\chi^2 = 0.01$, $p = 0.92$), suggesting

that the requirement for vernalization was determined by a single gene (*KVG1*) with complete dominance.

Fine mapping of *KVG1*

To fine map the *KVG1* gene, we designed several genetic markers in the target region on chromosome 2 based on the RNA-seq data (supplemental dataset 3). The 587 individuals from the segregating populations were genotyped, respectively using the far-end marker C02-1.301 and C02-3.671 to identify recombinants (supplemental dataset 3) (Fig. 1D). The recombinants were further genotyped using the other five markers in the target region (Fig. 1D). Consequently, *KVG1* was mapped between markers C02-2.036 and C02-2.117, spanning a region of approximately 80 kb (Fig. 1D).

FLC gene as the candidate gene of *KVG1*

The 80 kb candidate region contained 17 genes in cabbage reference genome (TO1000 V2.1) (Liu et al. 2014), but none of these genes are associated with flowering time according to their annotations (supplemental dataset 1) (Liu et al. 2014). We used these 17 genes to search for syntenic regions of other genomes in the species of Brassicaceae, including *B. rapa*, *B. napus,* and Arabidopsis. Interestingly, one *FLC* homolog was localized in the syntenic region of the genomes in *B. rapa*, *B. napus,* and Arabidopsis (Fig. 2).

Although no *FLC* homolog was detected in the reference genome of *B. oleracea*, we speculated that there might be *FLC* homolog in the candidate region of kale but not in that of Chinese kale. To test this speculation, a pair of primers located in the two flanking genes *Bo2g009720* and *Bo2g009710* were designed and were used to amplify PCR products from kale and Chinese kale (Fig. 3A). The results showed that the fragments of approximately 1 kb and 10 kb were amplified from Chinese kale and kale, respectively. The two amplified PCR products were sequenced, and 9,325 bp was found to be deleted from Chinese kale. In the deleted region, one 3,284-bp sequence encodes an FLC homolog. We referred to the *FLC* homolog as *BoFLC* hereafter. A pair of primers flanking the *BoFLC* (P1) were designed (Fig. 3A), and using this pair of primers, intact *BoFLC* was amplified from the kale parent (Fig. 3B).

Subsequently, we designed three primers located in *Bo2g00917*, *BoFLC* and *Bo2g009720*, respectively (Fig. 3A). These three primers were simultaneously used for PCR reaction to generate a codominant marker (Fig. 3C). This resultant codominant marker co-segregated with the vernalization trait in the 587 individuals. Therefore, we showed genetically that vernalization was not required in Chinese kale due to the deletion of a 9,325-bp fragment containing the *BoFLC* gene.
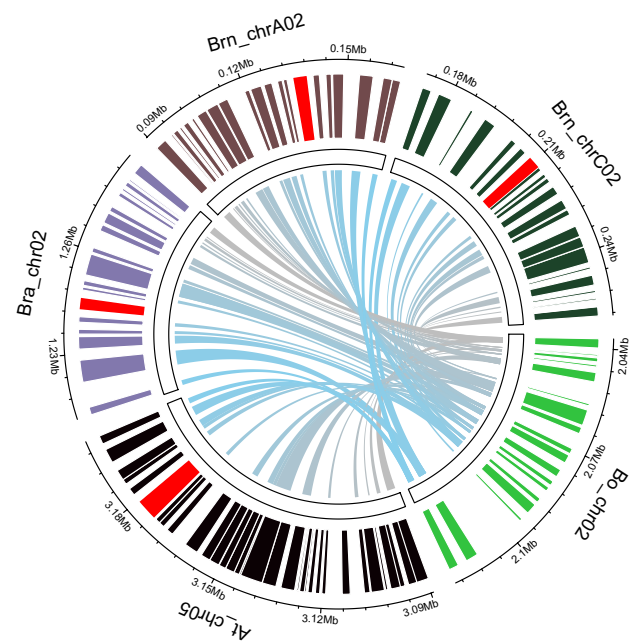

Fig. 2 Synteny analysis of the *KVG1* candidate region in different Brassica species and Arabidopsis. Each block on the outside layer presents the positions of genes in the synteny region, and the red blocks indicate *FLC* genes in the corresponding genomes. Bo_chr02, Bra_chr02, Brn_chrA02, Brn_chrC02, and At_chr05 represent the chromosomes of *B. oleracea*, *B. rapa*, *B. napus*, *B. napus*, and Arabidopsis, respectively. Linking lines inside the circle represent synteny pairs between two species

Sequence divergence among *FLC* homologs in *B. oleracea*

In addition to *BoFLC*, there existed four other *FLC* homologs in the genome of *B. oleracea*, including *Bo3g005470*, *Bo3g024250*, *Bo9g173370* and *Bo9g173400*. Phylogenetic analysis was carried out to explore the relationship between *FLC* homologs in the Brassicaceae species, including *B. oleracea*, *B. rapa*, *B. napus*, and Arabidopsis. The results revealed that the protein sequences of the orthologs/homologs were more closely related to each other than those of the paralogs within a genome (Fig. 4A), suggesting that the *FLC* homologs were duplicated before the speciation of Brassica species, which was consistent with the previous study (Schranz et al. 2002b).

Our manual annotation of the *FLC* homologs in the genome of *B. oleracea* revealed that a 1-bp insertion in the second exon of *Bo3g024250* led to frame shift and early termination. Since this 1-bp insertion also occurred in both the kale and Chinese kale parents, this gene was inferred to have no effect on flowering time variation between the two parents. Therefore, this gene was excluded from the subsequent investigation. The coding sequences of the other four *FLC* genes (excluding *Bo3g024250*) exhibited

Fig. 3 Deletion at the *KVG1* locus. **A** Positions of genes and primers in kale and Chinese kale. **B** Long PCR products amplified with P1 primers. lane 1, kale parent; lane 2, Chinese kale parent. **C** Codominant PCR products amplified with three primers act as a marker to detect the deletion of *BoFLC*. lane 1, the kale parent; lane 2, the Chinese kale parent; lane 3, Early flowering plant with homozygous genotype at the *KVG1* locus from the segregating population; lane 4, Late flowering plant with homozygous genotype at the *KVG1* locus from the segregating population; lane 5, Late flowering plant with heterozygous genotype in the segregating population
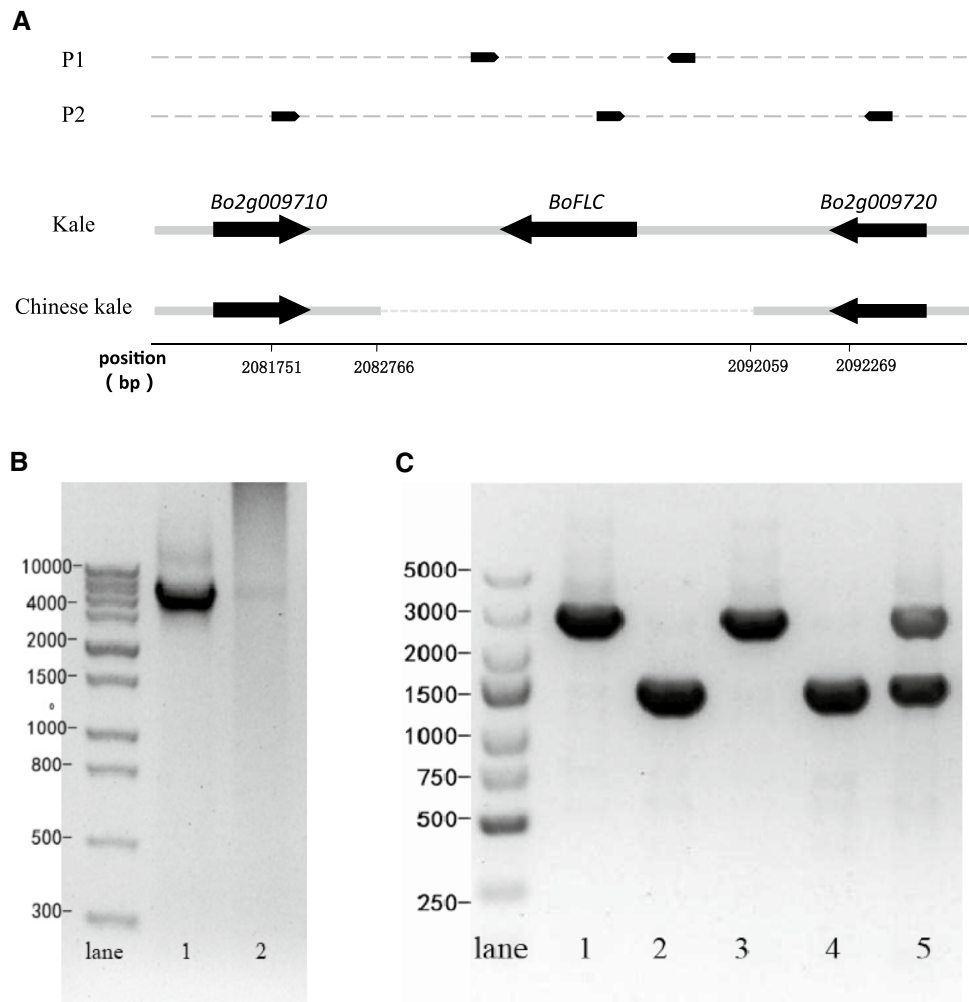
76.0–90.9% nucleotide identity and 84.2–97.8% amino acid identity. Protein sequences of these FLC homologs were highly conserved (Fig. 4B). In contrast to those of their highly conserved coding regions, the sequences of promoter regions of these four *FLC* homologs exhibited low nucleotide identity (less than 48.7%).

In order to investigate the genetic variations of the other three *FLC* genes, we designed PCR primers specific to each *FLC* homolog to amplify their sequences from the kale and Chinese kale parents (supplemental dataset 3). The *Bo9g173370* gene had identical coding sequences between the two parents, but this gene exhibited one 10-bp InDel in intron 4, and one 21-bp InDel, and four SNPs in the promoter region. The *Bo9g173400* gene was identical between the two parents. The *Bo3g005470* gene displayed one SNP in the coding sequence and six indels and multiple SNPs in introns. In summary, the presence/absence (P/A) of the *BoFLC* gene on chromosome 2 is the main difference in the *FLC* family between the kale and Chinese kale.

Gene expression analysis of *FLC* homologs in kale and Chinese Kale

Our data indicated that although there were five *FLC* homologs in the genomes of *B. oleracea,* vernalization was determined by *BoFLC* alone. We hypothesized that the lack of functional redundancy was caused by a low gene expression of other *FLC* genes. To test this hypothesis, we analyzed the expression of three *FLC* homologs (excluding *Bo3g024250*) in individuals of F_4 family (Fig. 5).

Interestingly, *Bo9g173370*, *Bo9g173400* and *Bo3g005470* exhibited no significant difference in expression level between early flowering individuals (without *BoFLC*) and nonflowering individuals (with intact *BoFLC*) (*t*-test; $p = 0.64$, 0.15, and 0.35). Furthermore, the expression levels of these three genes were much lower than that of *BoFLC* in late-flowering individuals. Consequently, the presence/absence of *BoFLC* determined the variation of total *FLC* expression levels, thus leading to variation on vernalization requirement. We conclude that the *FLC* homologs have no

Fig. 4 Sequence analysis of *FLC* homologs in Brassicaceae. **A** Phylogenetic tree of *FLC* homologs. Prefixes Bo, Br, Bna, and AT represent *FLC* homologs from *B. oleracea*, *B. rapa*, *B. napus,* and Arabidopsis, respectively. **B** Alignment of protein sequences of *FLC* homologs in *B. oleracea* and Arabidopsis. Shade indicates perfectly (black), highly (≥ 80%, dark grey), moderately (≥ 60%, grey) conserved residues
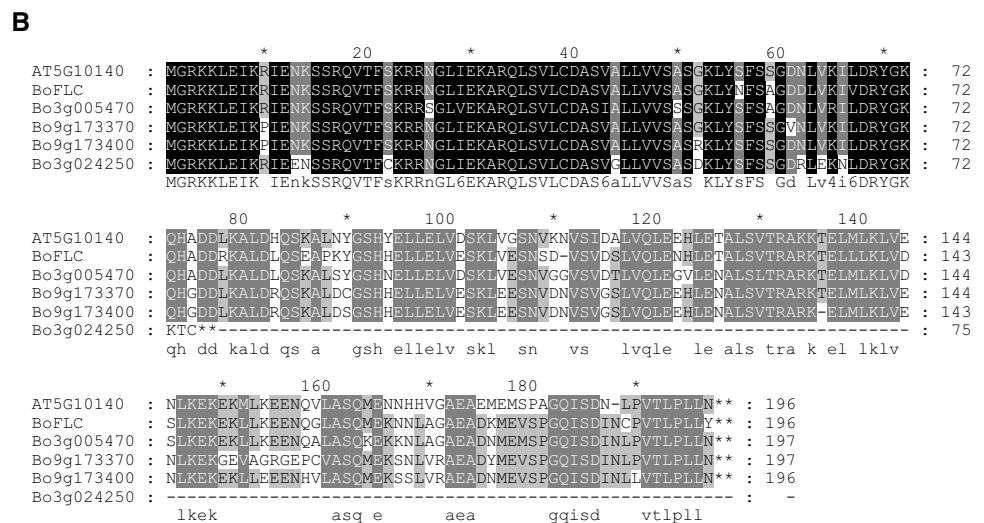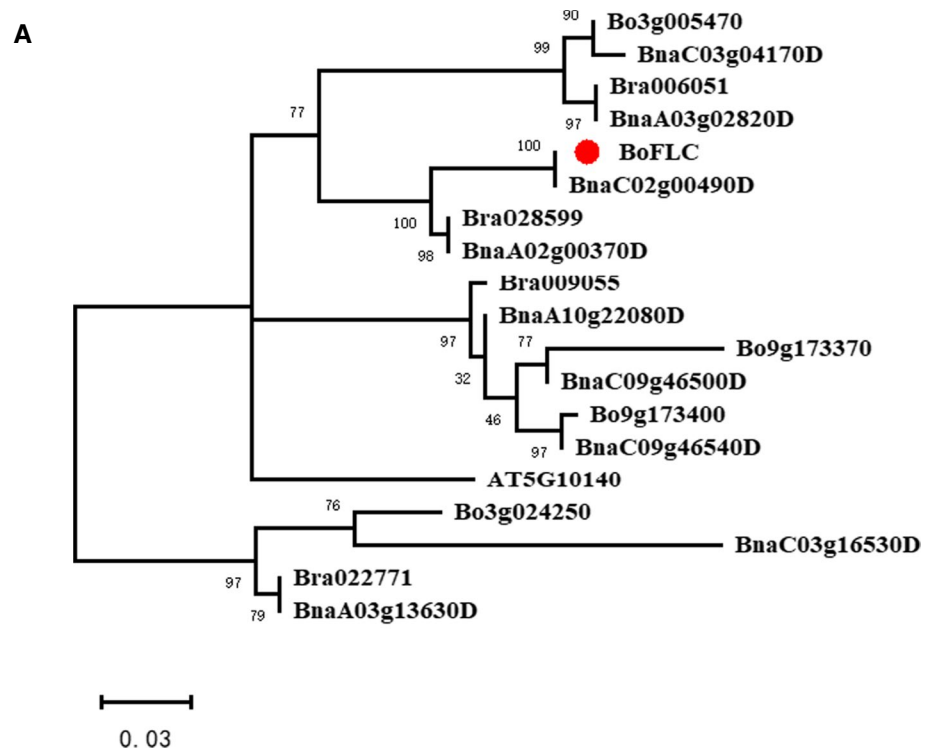
functional redundancy simply because other *FLC* genes have negligible expression.

Retainment of absence of *BoFLC* for early bolting in Chinese kale

A primer pair specific to the presence/absence polymorphism of *BoFLC* was used to genotype 42 cultivars of *B. oleracea*, including cabbage, broccoli, cauliflower, Chinese kale, kale, Brussels sprouts, and kohlrabi (supplemental dataset 2). Strikingly, all of the eight Chinese kales investigated in this study lost the *BoFLC* gene, while kale, cabbage, cauliflower, kohlrabi cultivars had the intact *BoFLC* gene (Fig. 6). This

is consistent with the fact that Chinese kale is the only non-vernalization crop in *B. oleracea*. Thus, we conclude that loss of *BoFLC* occurred during breeding program with the purpose of cultivating early bolting of Chinese kale since its stem was the main harvested and consumed parts.
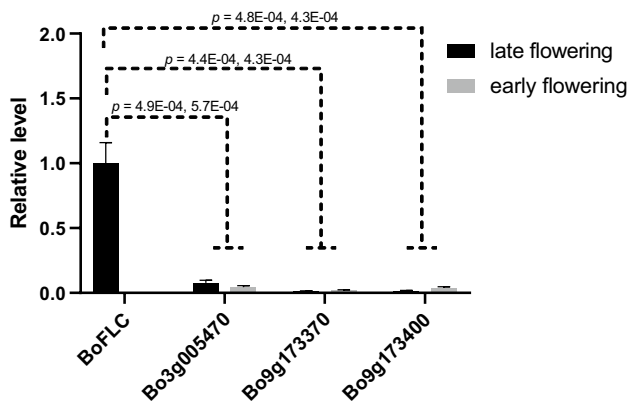
Fig. 5 qRT-PCR of *FLC* homologs in early flowering and late flowering individuals. The statistical significances were determined using Student's *t*-test. Error bars represent standard error
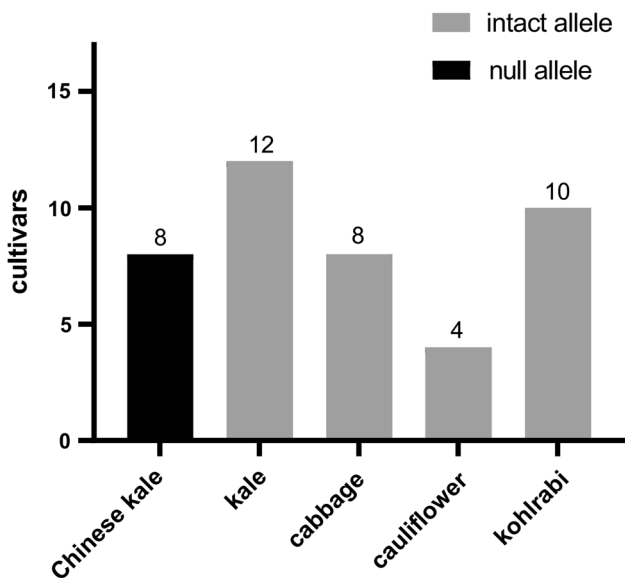

Fig. 6 Presence/absence of *BoFLC* in different cultivars of *B. oleracea*. The y-axis represents the number of cultivars

Discussion

BSA plus high-throughput sequencing facilitates cloning of genes controlling complex genetic traits

QTL mapping is commonly used to identify genetic regions controlling important traits by constructing recombinant inbred lines (RILs). RILs are usually established by several generations of inbreeding of an F_1 population (up to F_6 or F_7 populations) (Li et al. 2021; Qiao et al. 2021; Xu et al. 2021). Such an inbreeding process is labor- and time-consuming. In contrast, bulked segregant analysis is

a fast, economical, and efficient method for genetic analysis (Michelmore et al. 1991). High-throughput sequencing technology is now frequently used in combination with BSA to study complex genetic traits such as leaf colors and heading traits in lettuce (Su et al. 2020; Yu et al. 2020), fruit shape trait in wax gourd (Cheng et al. 2021), blossom-end rot trait in tomato (Topcu et al. 2021), grain weight in rice (Du et al. 2021), fruit spines in spinach (Liu et al. 2021). In this study, we employed BSA + RNA-seq method to identify the candidate gene *BoFLC*, which further confirmed that BSA + high-throughput sequencing could greatly facilitate the cloning of genes controlling complex genetic traits. In addition, when the candidate gene was absent in the reference genome of a certain species, syntenic region analysis of related species might contribute to the identification of the candidate gene.

FLC homologs play key roles in regulating flowering time in Brassica

One previous study has identified a QTL affecting flowering time in broccoli, which was syntenic to the region of *BoFLC*. The candidate gene controlling flowering time in broccoli was predicted to be *Bol024659*, a homolog of *GRF6* in Arabidopsis (Shu et al. 2018). However, the *Bol024659* gene was located outside the 80 kb candidate region of *KVG1* in this study, therefore, this gene was excluded as the candidate gene for *KVG1*. Our data showed that the absence of *BoFLC* was responsible for non-vernalization in Chinese kale.

Previous QTL analyses of flowering time in Brassica have revealed that the *FLC* family plays a key role in regulating flowering time in *B. napus*, *B. oleracea*, and *B. rapa*. Specifically, in *B. rapa*, a naturally occurring deletion spanning exon 4 and intron 4 in *BrFLC2*, and a G/A polymorphism at the 5`splice site in intron 6 of *BrFLC1* contribute to flowering time variations (Wu et al. 2012; Yuan et al. 2009). In *B. napus*, a *Tourist*-like MITE insertion into the upstream region of the *BnFLC.A10* gene affects flowering time in rapeseed (Hou et al. 2012). In *B. oleracea*, the *BoFLC.C2* was co-localized with one QTL regulating flowering time in broccoli (Irwin et al. 2016; Okazaki et al. 2007). The above studies indicate that the polymorphisms of *FLC* homologs in Brassica affect flowering time, but those polymorphisms did not change the vernalization requirement. Our current study reveals that the absence of *BoFLC* can lead to Chinese kale flowering without vernalization. Although the *BoFLC* is the major component for controlling flowering time (vernalization) variation, the effects of other *BoFLC* homologs could not be excluded in *B. oleracea* species.

Total expression level of *FLC* homologs controls flowering time

The role of *FLC* homologs in regulating flowering time is similar across Brassicaceae species (Leijten et al. 2018). The *B. oleracea* genome has five *FLC* homologs. The retention of *FLC* paralogs contributes to the maintenance of stoichiometric expression balance, and over-retention of the *FLC* paralogs suggests dosage-sensitivity of *FLC* family (Maere et al. 2005). If selection pressure to keep balance acts on the total gene expression level of a functionally similar gene family, the expression level of individual paralogs will fluctuate, and this effect caused by this fluctuation is compensated for by other paralogs (Birchler and Veitia 2010, 2012; Thompson et al. 2016). Total expression of all the *BnaFLC* paralogs rather than the expression of individual *BnaFLC* paralogs determine vernalization requirement in *B. napus* (Calderwood et al. 2021). In the current study, we discovered that the expression of *BoFLC* was much higher than that of its paralogs in *B. oleracea.* Thus, the *BoFLC* was decisive for the vernalization requirement. Polymorphisms were also detected in other *FLC* paralogs, and they might also affect flowering time variation but not vernalization requirement.

Presence/absence of *BoFLC* can be used in future breeding of *B. oleracea*

B. oleracea includes several cultivars such as cabbage, broccoli, cauliflower, Chinese kale, kale, Brussels sprouts, and kohlrabi. The main consumed part of kohlrabi is swollen stem, and bolting stem for Chinese kale. Except for Chinese kale and kohlrabi, leaves are the main edible part of other cultivars. The loss of *BoFLC* is only detected in Chinese kale. Early flowering with no vernalization requirement is a favorable trait for Chinese kale breeding. We speculated that the 9,325-bp deletion fragment containing the *BoFLC* gene was selected for the early flowering of Chinese kale. The null allele of *BoFLC* could be introduced into other *B. oleracea* sub-species such as cabbage to shorten the reproduction time of inbred lines. If an inbred line with the null allele is crossed with an inbred line with the intact allele of *BoFLC,* its vernalization requirement can be recovered in the resulting hybrid cultivar.

Authors' contribution statement QT and YJ performed the molecular experiments; YJ and WZ did the bioinformatic analysis. HK designed the experiment; YJ wrote the manuscript with the help of HK. GA, RT, and CY gave the critical review of this manuscript.

Declarations

References

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol 186:54–62

Birchler JA, Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci USA 109:14746–14753

Blumel M, Dally N, Jung C (2015) Flowering time regulation in crops-what did we learn from Arabidopsis? Curr Opin Biotechnol 32:121–129

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60

Burn JE, Smyth DR, Peacock WJ, Dennis ES (1993) Genes conferring late flowering in arabidopsis-Thaliana. Genetica 90:147–155

Calderwood A, Lloyd A, Hepworth J, Tudor EH, Jones DM, Woodhouse S, Bilham L, Chinoy C, Williams K, Corke F, Doonan JH, Ostergaard L, Irwin JA, Wells R, Morris RJ (2021) Total *FLC* transcript dynamics from divergent paralogue expression explains flowering diversity in Brassica napus. New Phytol 229:3534–3548

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VH, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CH, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker

P (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345:950–953

Cheng Z, Liu Z, Xu Y, Ma L, Chen J, Gou J, Su L, Wu W, Chen Y, Yu W, Wang P (2021) Fine mapping and identification of the candidate gene BFS for fruit shape in wax gourd (Benincasa hispida). Theor Appl Genet. https://doi.org/10.1007/s00122-021-03942-8. Online ahead of print. PMID: 34480584

Clarke JH, Dean C (1994) Mapping *FRI*, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. Mol Gen Genet 242:81–89

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES (2011) *FLOWERING LOCUS C* (*FLC*) regulates development pathways throughout the life cycle of Arabidopsis. Proc Natl Acad Sci U S A 108:6680–6685

Du Z, Huang Z, Li J, Bao J, Tu H, Zeng C, Wu Z, Fu H, Xu J, Zhou D, Zhu C, Fu J, He H (2021) qTGW12a, a naturally varying QTL, regulates grain weight in rice. Theor Appl Genet 134:2767–2776

Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E, Knee E, Lambrecht M, Miller N, Mueller LA, Mundodi S, Reiser L, Rhee SY, Scholl R, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2002) TAIR: a resource for integrated Arabidopsis data. Funct Integr Genomics 2:239–253

Hou J, Long Y, Raman H, Zou X, Wang J, Dai S, Xiao Q, Li C, Fan L, Liu B, Meng J (2012) A Tourist-like MITE insertion in the upstream region of the *BnFLC.A10* gene is associated with vernalization requirement in rapeseed (Brassica napus L.). BMC Plant Biol 12:238

Irwin JA, Soumpourou E, Lister C, Ligthart JD, Kennedy S, Dean C (2016) Nucleotide polymorphism affecting *FLC* expression underpins heading date variation in horticultural brassicas. Plant J 87:597–605

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science 290:344–347

Koornneef M, Vrles HB, Hanhart C, Soppe W, Peeters T (1994) The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild-type. Plant J 6:911–919

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinf (oxf Engl) 23:2947–2948

Lei J, Chen G, Chen C, Cao B (2017) Germplasm diversity of Chinese Kale in China. Hortic Plant J 3:101–104

Leijten W, Koes R, Roobeek I, Frugis G (2018) Translating flowering time from arabidopsis thaliana to brassicaceae and asteraceae crop species. Plants (Basel) 7

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinf (oxf Engl) 25:2078–2079

Li Y, Ye H, Song L, Vuong TD, Song Q, Zhao L, Shannon JG, Li Y, Nguyen HT (2021) Identification and characterization of novel QTL conferring internal detoxification of aluminium in soybean. J Exp Bot 72:4993–5009

Lin SI, Wang JG, Poon SY, Su CL, Wang SS, Chiou TJ (2005) Differential regulation of *FLOWERING LOCUS C* expression by vernalization in cabbage and Arabidopsis. Plant Physiol 137:1037–1048

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang TJ, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park BS, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee TH, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930

Liu Z, Lu T, Feng C, Zhang H, Xu Z, Correll JC, Qian W (2021) Fine mapping and molecular marker development of the Fs gene controlling fruit spines in spinach (Spinacia oleracea L.). Theor Appl Genet 134:1319–1328

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (san Diego, Calif) 25:402–408

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA 102:5454–5459

Mempel TR, Henrickson SE, Von Andrian UH (2004) T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. Nature 427:154–159

Michaels SC, Amasino RM (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11:949–956

Michelmore R, W., Paran I, Kesseli RV, (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci 88:9828–9832

Okazaki K, Sakamoto K, Kikuchi R, Saito A, Togashi E, Kuginuki Y, Matsumoto S, Hirai M (2007) Mapping and characterization of *FLC* homologs and QTL analysis of flowering time in Brassica oleracea. Theor Appl Genet 114:595–608

Osborn TC, Kole C, Parkin IA, Sharpe AG, Kuiper M, Lydiate DJ, Trick M (1997) Comparison of flowering time genes in Brassica rapa, B. napus and Arabidopsis Thaliana. Genetics 146:1123–1129

Pires JC, Zhao J, Schranz ME, Leon EJ, Pablo AQ, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae). Biol J Lin Soc 82:675–688

Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Report 15:8–15

Qian H, Liu T, Deng M, Miao H, Cai C, Shen W, Wang Q (2016) Effects of light quality on main health-promoting compounds and antioxidant capacity of Chinese kale sprouts. Food Chem 196:1232–1238

Qiao J, Jiang H, Lin Y, Shang L, Wang M, Li D, Fu X, Geisler M, Qi Y, Gao Z, Qian Q (2021) A Novel miR167a-OsARF6-OsAUX3 Module regulates grain length and weight in rice. Mole plant

Razi H, Howell EC, Newbury HJ, Kearsey MJ (2008) Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in Brassica oleracea? Theor Appl Genet 116:179–192

Ridge S, Brown PH, Hecht V, Driessen RG, Weller JL (2015) The role of *BoFLC2* in cauliflower (Brassica oleracea var. botrytis L.) reproductive development. J Exp Bot 66:125–135

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon R (2017) Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci Rep 7:41845

Schranz ME, Quijada P, Sung S, Lukens L, Amasino R, Osborn TC (2002a) Characterization and effects of the replicated flowering time gene FLC in Brassica rapa. Genet Soc Am 162:1457–1468

Schranz ME, Quijada P, Sung SB, Lukens L, Amasino R, Osborn TC (2002b) Characterization and effects of the replicated flowering time gene *FLC* in Brassica rapa. Genetics 162:1457–1468

Searle I, He Y, Turck F, Vincent C, Fornara F, Krober S, Amasino RA, Coupland G (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes Dev 20:898–912

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (1999) The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell 11:445–458

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). Proc Natl Acad Sci USA 97:3753–3758

Shu J, Liu Y, Zhang L, Li Z, Fang Z, Yang L, Zhuang M, Zhang Y, Lv H (2018) QTL-seq for rapid identification of candidate genes for flowering time in broccoli x cabbage. Theor Appl Genet 131:917–928

Song J, Irwin J, Dean C (2013) Remembering the prolonged cold of winter. Curr Biol 23:R807-811

Su W, Tao R, Liu W, Yu C, Yue Z, He S, Lavelle D, Zhang W, Zhang L, An G, Zhang Y, Hu Q, Larkin RM, Michelmore RW, Kuang H, Chen J (2020) Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. Plant Biotechnol J 18:479–490

Sun B, Liu N, Zhao Y, Yan H, Wang Q (2011) Variation of glucosinolates in three edible parts of Chinese kale (Brassica alboglabra Bailey) varieties. Food Chem 124:941–947

Sun B, Yan H, Zhang F, Wang Q (2012) Effects of plant hormones on main health-promoting compounds and antioxidant capacity of Chinese kale. Food Res Int 48:359–366

Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, Peacock WJ (2001) Control of flowering time by FLC orthologues in Brassica napus. Plant J 28:545–553

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S, Terauchi R (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183

Thompson A, Zakon HH, Kirkpatrick M (2016) Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. Genetics 202:765–774

Topcu Y, Sapkota M, Illa-Berenguer E, Nambeesan SU, van der Knaap E (2021) Identification of blossom-end rot loci using joint QTL-seq and linkage-based QTL mapping in tomato. Theor Appl Genet 134:2931–2945

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3–new capabilities and interfaces. Nucl Acids Res 40:e115

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011) The genome of the mesopolyploid crop species Brassica rapa. Nat Genet 43:1035–1039

Wei J, Miao H, Wang Q (2011) Effect of glucose on glucosinolates, antioxidants and metabolic enzymes in Brassica sprouts. Sci Hortic 129:535–540

Wu J, Wei K, Cheng F, Li S, Wang Q, Zhao J, Bonnema G, Wang X (2012) A naturally occurring InDel variation in *BraA.FLC.b* (BrFLC2) associated with flowering time variation in Brassica rapa. BMC Plant Biol 12:151

Wu S, Lei J, Chen G, Chen H, Cao B, Chen C (2017) De novo transcriptome assembly of chinese kale and global expression analysis of genes involved in glucosinolate metabolism in multiple tissues. Front Plant Sci 8:92

Xu Z, Wang F, Zhou Z, Meng Q, Chen Y, Han X, Tie S, Liu C, Hao Z, Li M, Zhang D, Han J, Wang Z, Li X, Weng J (2021) Identification and Fine-Mapping of a Novel QTL qMrdd2 Conferring Resistance to Maize Rough Dwarf Disease. Plant Dis. https://doi.org/10.1094/PDIS-03-20-0495-RE. Online ahead of print, PMID: 34132596

Yu C, Yan C, Liu Y, Liu Y, Jia Y, Lavelle D, An G, Zhang W, Zhang L, Han R, Larkin RM, Chen J, Michelmore RW, Kuang H (2020) Upregulation of a *KN1* homolog by transposon insertion promotes leafy head development in lettuce. Proc Natl Acad Sci USA 117:33668–33678

Yuan YX, Wu J, Sun RF, Zhang XW, Xu DH, Bonnema G, Wang XW (2009) A naturally occurring splicing site mutation in the Brassica rapa *FLC1* gene is associated with variation in flowering time. J Exp Bot 60:1299–1308

Zhao J, Kulkarni V, Liu N, Del Carpio DP, Bucher J, Bonnema G (2010) *BrFLC2* (*FLOWERING LOCUS C*) as a candidate gene for a vernalization response QTL in Brassica rapa. J Exp Bot 61:1817–1825