Theor Appl Genet (2008) 116:179–192
DOI 10.1007/s00122-007-0657-3

ORIGINAL PAPER

# Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*?

**H. Razi · E. C. Howell · H. J. Newbury · M. J. Kearsey**

Received: 5 February 2007 / Accepted: 21 September 2007 / Published online: 16 October 2007


**Abstract** Previous locations of flowering time (FT) QTL in several *Brassica* species, coupled with *Arabidopsis* synteny, suggest that orthologues of the genes *FLC, FY* or *CONSTANS* might be the candidates. We focused on *FLC*, and cloned paralogous copies in *Brassica oleracea*, obtained their genomic DNA sequences, and confirmed their locations relative to those of known FT-QTL by genetical mapping. They varied in total length mainly due to the variable size of the first and last introns. A high level of identity was observed among *Brassica FLC* genes at the amino acid level but non-synonymous differences were present. Comparative analysis of the promoter and intragenic regions of *BoFLC* paralogues with *Arabidopsis FLC* revealed extensive differences in overall structure and organisation but showed high conservation within those segments known to be essential in regulating *FLC* expression. Four *B. oleracea FLC* copies (*BoFLC1*, *BoFLC3, BoFLC4* and *BoFLC5*) were located to their respective linkage groups based on allelic sequence variation in lines from a doubled haploid population. All except *BoFLC4* were within the confidence intervals of known FT-QTL. Sequence data indicated that relevant non-synonymous polymorphisms were present between parents A12DHd and GDDH33 for *BoFLC* genes. However, *BoFLC* alleles segregated independently of FT in backcrosses while the study provided evidence that *BoFLC4* and *BoFLC5* contain premature stop codons and so could not contribute to flowering time variation. Therefore, there is strong evidence against any of the 4 *BoFLC* being FT-QTL candidates in this population.

Communicated by A. Bervillé.

H. Razi · E. C. Howell · H. J. Newbury · M. J. Kearsey
School of Biosciences, The University of Birmingham,
Birmingham B15 2TT, UK

H. Razi (✉)
Department of Crop Production and Plant Breeding,
College of Agriculture, Shiraz University,
Shiraz 71441-65186, Iran
e-mail: razi@shirazu.ac.ir

## Introduction

A key priority of contemporary crop genetics and plant breeding is to identify the genes underlying quantitative trait loci (QTL); one obvious initial route to this end is candidate gene analysis. Flowering (FT) time is an excellent system to explore the candidate gene approach because, on the one hand, numerous studies have explored populations to locate FT-QTL, while on the other we have extensive detailed knowledge, largely from *Arabidopsis*, of the key genes involved in controlling FT. It is a trait of importance in plant breeding, which has a high heritability in most species and responds readily to artificial and natural selection.

Numerous studies in different *Brassica* species have identified FT-QTL in very similar paralogous chromosomal regions that are syntenous to a region at the top of *Arabidopsis* chromosome 5 (At5) (Lagercrantz et al. 1996; Osborn et al. 1997; Bohuon et al. 1998; Rae et al. 1999; Axelsson et al. 2001; Parkin et al. 2002; Schranz et al. 2002; Okazaki et al. 2007). This is a region where several well characterised flowering time genes such as *FLC*, *FY* and *CONSTANS (CO)* are located. The *CO* homologues identified on linkage groups 2 and 8 of *B. nigra* were coincident with the genes controlling flowering time (Lagercrantz et al. 1996). *CO* orthologues appeared to be close to the most likely QTL positions in eight out of nine QTL controlling flowering

time studied in different *Brassica* species (*B. rapa*, *B. oleracea* and *B. juncea*), whereas *FLC* orthologues were mapped farther away (Axelsson et al. 2001). Of six FT-QTL found using a doubled haploid population in *B. oleracea*, three were mapped to genomic regions on linkage groups O2, O3 and O9 syntenic with the *CO* contig at the top of At5 (Bohuon et al. 1998). These were confirmed and more FT-QTL were later identified using substitution lines derived from the same parental lines (Rae et al. 1999). Okazaki et al. (2007) found *BoFLC2* as a putative candidate gene for a large effect FT-QTL on linkage group O2. They did not detect any FT-QTL in the regions in which other *BoFLC* copies were mapped. In *B. napus*, a genomic region containing a FT-QTL called *VFN2* showed strong synteny with the top of At5 (Osborn et al. 1997). It was reported that *VFR2*, which is one of the major QTL controlling vernalization-responsive flowering time in *B. rapa* (Osborn et al. 1997), might be homologous to *FLC* (Kole et al. 2001). Further investigation revealed that *BrFLC1* co-segregates with *VFR2* (Schranz et al. 2002). *BrFLC2* and *BrFLC5* were also proposed as possible candidates for two FT-QTL identified on chromosomes R2 and R3 in *B. rapa* (Schranz et al. 2002).

Genetic redundancy, as occurs in diploid and amphidiploid *Brassica* species, is a potential source of novel genetic, and hence phenotypic, variation. It is assumed that the diploid *Brassicas* are derived from an ancient hexaploid ~10 My BP (Lagercrantz and Lydiate 1996; Lagercrantz 1998; Parkin et al. 2005) because of extensive triplication across their genomes (Cavell et al. 1998; O'Neil and Bancroft 2000; Parkin et al. 2002; Yang et al. 2005). Replicated copies of genes may continue to function similarly to the ancestral gene to jointly influence the target trait, they may diverge and gain new functions to broaden phenotypic and genotypic effects, or one or more copies may lose function (Lagercrantz and Axelsson 2000; Wendel 2000). High levels of genome replication in *Brassica* species also complicate the characterisation and evaluation of individual genes in a genome-wide context.

Comparative microsynteny studies disclosed a high rate of conservation between *Arabidopsis* and *Brassica* sequences particularly in coding regions (Roberts et al. 1998; Quiros et al. 2001; Schranz et al. 2002; Suzuki et al. 2003; Ayele et al. 2005), although sequence alterations and disruptions of gene content were higher than predicted by comparative genetic mapping experiments (O'Neil and Bancroft 2000; Quiros et al. 2001). Such a high similarity in structure and function makes it possible to explore major physiological and developmental processes in *Brassicas* using *Arabidopsis* information.

While more than 80 genes that control flowering time in *Arabidopsis* (Levy and Dean 1998) have been identified, natural variation for FT mainly results from allelic differences at two regulatory genes, *FRI* and *FLC*, which are central genes in vernalization requirement and response (Gazzani et al. 2003; Shindo et al. 2005). The genetic and molecular basis of vernalization has been comprehensively studied in *Arabidopsis* (Michaels and Amasino 2000; Sheldon et al. 2000a; Henderson et al. 2003; Sung and Amasino 2005).

In this paper we will consider the candidate gene *FLC*, which encodes a MADS domain transcription factor which represses flowering in a dosage-dependent manner (Michaels and Amasino 1999; Sheldon et al. 1999). A high level of *FLC* expression is detected in vernalization-responsive late flowering accessions while even higher levels occur in over-expressing transgenic plants. Conversely, early flowering ecotypes show low levels of *FLC* transcript and protein (Sheldon et al. 1999, 2000b; Michaels and Amasino 2000). *FLC* activity is exerted through down-regulation of floral pathway integrator genes, *FT* and *SOC1* (Hepworth et al. 2002; Michaels et al. 2005). *FLC* expression is mainly up regulated by *FRI* synergistically to cause late flowering (Michaels and Amasino 1999; Michaels and Amasino 2000). Conversely, vernalization strongly decreases *FLC*-mRNA level and so shortens flowering time (Michaels and Amasino 1999; Sheldon et al. 1999). Different regions within the promoter and intron 1 are required to regulate *Arabidopsis FLC* activities, including non-vernalized expression, initial down-regulation and maintenance of repression induced by vernalization (Sheldon et al. 2002; Michaels et al. 2003; Bastow et al. 2004).

A number of *Arabidopsis FLC* orthologues have been cloned in *Brassica* crop species (Tadege et al. 2001; Schranz et al. 2002; Martynov and Khavkin 2004; Li et al. 2005; Lin et al. 2005; Yang et al. 2006; Okazaki et al. 2007). Phylogenetic reconstruction supported orthology between *Brassica FLC* genes and *Arabidopsis FLC* (Tadege et al. 2001; Schranz et al. 2002; Li et al. 2005) because they are more similar to *Arabidopsis FLC* than *mads affecting flowering* (*MAF*) genes, which have been shown to be the closest gene family to *FLC* in *Arabidopsis* (Ratcliffe et al. 2003).

Three *FLC* genes (*BoFLC1*, *BoFLC3* and *BoFLC5*) were isolated and partially sequenced in a rapid cycling line of *B. oleracea* (Schranz et al. 2002). A tandem duplication of *BoFLC1* was later identified (Salathia 2003). *BoFLC3* and *BoFLC5* were shown to be on linkage group O3 (Pires et al. 2004; Okazaki et al. 2007), while *BoFLC1* was assigned to linkage group O9 (Salathia 2003; Pires et al. 2004). In contrast, Okazaki et al. (2007) mapped *BoFLC1* to the top of linkage group O2. Also, two fully sequenced *FLC* copies, *BoFLC3-2* and *BoFLC4-1*, were reported in *B. oleracea* var. *capitata* (Lin et al. 2005). More recently, Okazaki et al. (2007) identified a copy of *FLC* in *B. oleracea* (*BoFLC2*)

mapped on O2, which showed very high homology (98%) to *BoFLC4-1*. Despite great similarity with *Arabidopsis FLC* in coding sequences, variation in promoter and intronic regions were presumed to cause different regulatory mechanisms in response to vernalization between *Brassica* and *Arabidopsis* (Lin et al. 2005).

In the present study, three fully sequenced *B. oleracea FLC* genes including promoter regions are reported and their genomic organisation and polymorphism in a QTL mapping population is characterised. Intragenic and promoter regions of *BoFLC* genes are compared to *Arabidopsis FLC* and phylogenetic relationships among *Brassica FLCs* are analysed. We also compare the locations of known FT-QTL in *Brassica oleracea* with the position of four replicated *FLC* copies using genetical mapping. The results provide evidence to assess whether the *BoFLC* genes are possible candidates for three QTL identified in the mapping population or whether the other linked genes are responsible.

## Materials and methods

### Plant materials

The source material was derived from two parental lines, A12DHd and GDDH33, obtained from separate microspore-derived doubled haploid (DH) lines of *B. oleracea*. A12DHd is a rapid-cycling and self-compatible line derived from *B. oleracea* var. *alboglabra*. GDDH33 came from a commercial $F_1$ hybrid Calabrese variety, Green Duke (*B. oleracea* var. *italica*) (Bohuon et al. 1998) and is a late-flowering and self-incompatible genotype. For simplicity, they will be referred to as A12 and GD, respectively.

We used selected DH lines from the A12 × GD population, which had been genotyped for ~430 molecular markers (Sebastian et al. 2000; *Brassica* Genome Gateway website: http://brassica.bbsrc.ac.uk/) to locate *FLC* genes. A set of DH lines was used to assign *FLC* copies to their appropriate linkage groups. Subsequently, each *FLC* was positioned more precisely using additional DH lines with crossovers at different points within these regions.

Three *B. oleracea* part-chromosome substitution lines (SL128, SL133 and SL175) were selected from a sample of 79 substitution lines produced from A12 and GD (Ramsay et al. 1996). These lines flowered significantly later than the early flowering parent (A12) due to their single introgressed region of GD-DNA (Rae et al. 1999). These regions overlapped the confidence intervals of three FT-QTL identified in the A12 × GD DH mapping population (Bohuon et al. 1998; Rae et al. 1999). The substituted regions of these lines were predicted to contain the locations of three *Brassica FLC* copies (*BoFLC1*, *BoFLC3* and *BoFLC5*) on linkage groups O9, O3 and O3, respectively. $F_1$ plants, derived from crosses between the three substitution lines and A12, were backcrossed to A12 and the three backcross families were sown along with the recurrent parent (A12) in a glasshouse on May 2004. The latest flowering individuals in each backcross family were self-pollinated to produce $BC_1S_1$ seed. Sets of $BC_1S_1$ individuals, 169, 177 and 192 derived from SL133, SL128 and SL175, respectively, were sown alongside 15 replicates of A12 in mid-February 2005 in a glasshouse trial to evaluate flowering time in a fully randomised trial with guards. Flowering time was recorded as the number of days from sowing to the appearance of the first flower opening on each individual.

### Sub-clone library; construction and screening

Three previously identified BAC clones, which carried *FLC* genes (Razi 2006) were used to construct a sub-clone library. The BAC clones 032J18, 043I02 and 019H24 contained *BoFLC1*, *BoFLC3* and *BoFLC5*, respectively, and were from the *B. oleracea JBo*BAC library constructed from the A12 genome (O'Neil and Bancroft 2000). The clones were grown on LB agar plates containing 12.5 μg/ml kanamycin. Single colonies were picked and grown overnight in a large volume of LB broth containing kanamycin. DNA was extracted on a large scale using QIAGEN Large-Construct kit (Qiagen, Valencia, CA, USA) to obtain ~20 μg DNA. The TOPO Shotgun sub-cloning kit (Invitrogen, Paisley, UK) was employed to construct sub-clone libraries, following the manufacturer's recommended protocol. PCR screening of the three resulting libraries was performed in several rounds in order to build up an overlapping series of inserts and eventually complete the sequence of all *FLC* genes. Primers corresponding to the sequences flanking the inserts in the vector, T3 (5′-ATTAACCCTCACTAAAGGGA-3′) and T7 (5′-AATACGACTCACTATAGGG-3′), were exploited accompanied with primers designed from *FLC* genes to amplify and then sequence the inserts of positive clones. The sequences of all primers are given in Table 1.

PCR was carried out as follows: 2 min 30 s initial denaturation at 95°C followed by 30 cycles of amplification including 30 s for DNA denaturation at 95°C, 30 s for annealing primers at 56°C and 1 min for elongation at 72°C. The last step was 2 min 30 s for final extension at 72°C. PCR products were separated on 0.8% agarose gel, purified and quantified prior to sequencing. PCR products were sequenced by ABI 3700 sequencer (Applied Biosystems Inc., Foster City, CA, USA) using the dideoxy chain termination method. Sequencing was performed using the Big Dye Terminator labelling mix following the manufacturer's instructions.

**Table 1** Forward and reverse primers used to amplify different *FLCs* in *Brassica oleracea* (Bo)

| Primer | Sequence |
|---|---|
| *FLC1 primers* | |
| Bo FLC1 F1 | 5′-CCTAGACGGGTCCGAATCTGGGAC-3′ |
| Bo FLC1 F2 | 5′-GTCTCTAATTGTCTTCTGTGCCC-3′ |
| Bo FLC1 F3 | 5′-GGTGTTAGAATGTATTGGCATGCCC-3′ |
| Bo FLC1 F4 | 5′-CTTGCTCAAGGGTCCAGTGGTG-3′ |
| Bo FLC1 F5 | 5′-CATCCGTCTATTCAAACGTCCG-3′ |
| Bo FLC1 R1 | 5′-GTCCCAGATTCGGACCCGTCTAGG-3′ |
| Bo FLC1 R2 | 5′-GGTTGTCTCATGTATCTAGCAAC-3′ |
| Bo FLC1 R3 | 5′-CGGTACGGGTTCGGTTCGGATTTC-3′ |
| Bo FLC1 R4 | 5′-GAGTGCGAAACTAAACGCTTGGG-3′ |
| Bo FLC1 R5 | 5′-GGCGTAGAGGTAATCCATAGAAGC-3′ |
| Bo FLC1 R6 | 5′-GGTTGTGCATGAGGATCCATCA-3′ |
| Bo FLC1 R7 | 5’-GCAGTGGGAGCGTTACCGGAAG-3’ |
| Bo FLC1 R8 | 5′-GAGCTGAAGATACATGGGAGCGAG-3′ |
| *FLC3 primers* | |
| Bo FLC3 F1 | 5′-GGTACACGTGGCTGTCTTCTCGTC-3′ |
| Bo FLC3 F2 | 5′-GAGGTATTGCATTGTTGGTCCACC-3′ |
| Bo FLC3 F3 | 5′-GTGCCGGTGTTCATTCAAATTTGG-3′ |
| Bo FLC3 R1 | 5′-CGGGTACCCGAAATATTTCGGTTC-3′ |
| Bo FLC3 R2 | 5′-GTCAATAGCTGGACAATGTCGTAC-3′ |
| Bo FLC3 R3 | 5′-CCAGGGCTTTAAGATCATCAGC-3′ |
| Bo FLC3 R4 | 5′-GACTGAAGATCCTGTCCACGGAG-3′ |
| Bo FLC3 R5 | 5′-CAAGAAGTGCTTATCGGCTTTTGC-3′ |
| Bo FLC3 R6 | 5′-CTCCATATTATCAGCTTCGGCTCCC-3′ |
| Bo FLC3 R7 | 5′-GGAGTACACACAATCTCTCAGCC-3′ |
| *FLC4 primers* | |
| Bo FLC4 F1 | 5′-CTCCTCTTCAGCCTGGTCAAGGTC-3′ |
| Bo FLC4 F2 | 5′-GACAGGATCTTCAGTCAGAAGCTCC-3′ |
| Bo FLC4 F3 | 5′-CGAATGTATGCCACATTGTGCAGC-3′ |
| Bo FLC4 R1 | 5′-CTAACAAAAACGCCCTTCTCGGC-3′ |
| Bo FLC4 R2 | 5′-GGGCATCTCCGTCCCAACTCCAT-3′ |
| Bo FLC4 R3 | 5′-GGAGCTTCTGACTGAAGATCCTGTC-3′ |
| *FLC5 primers* | |
| Bo FLC5 F1 | 5′-GCGGTGCACGTGGCTGTCTTGTCG-3′ |
| Bo FLC5 F2 | 5′-GAGAGATCTCAGAATATACTCTCG-3′ |
| Bo FLC5 F3 | 5′-CTGAATGCTAGGTTCAGCCTTGG-3′ |
| Bo FLC5 F4 | 5′-GGTAGATTCCAGTGGTGTCTTC-3′ |
| Bo FLC5 F5 | 5′-ATACACTGGTCCTTTACCGCCTC-3′ |
| Bo FLC5 F6 | 5′-GCTTTCCTAGCTAGTTCAGCCAGG-3′ |
| Bo FLC5 R1 | 5′-CGGAGGAGAAGCTGTAGAGCTTG-3′ |
| Bo FLC5 R2 | 5′-GATTCGCCGGGTAAATCTAAGTGTC-3′ |
| Bo FLC5 R3 | 5′-CGCGAAGAGACAGCCAACGGTATC-3′ |
| Bo FLC5 R4 | 5′-GGTATCAGAGGGTTCTAGCGATCC-3′ |
| Bo FLC5 R5 | 5′-GACTAATGGAACTCGGCACTAAC-3′ |
| Bo FLC5 R6 | 5′-CAACTGATGCACATTACGTGCTGC-3′ |
| Bo FLC5 R7 | 5′-CAGGGCGTGTGTTGCTGCACTTCC-3′ |
| Bo FLC5 R8 | 5′-CCAGGGCATTGAGATCATCAGC-3′ |
| Bo FLC5 R9 | 5′-CATTTGTAATGAAAGGAGGAGAGC-3′ |
| Bo FLC5 R10 | 5′-CCTGGCTGAACTAGCTAGGAAAGC-3′ |
| Bo FLC5 R11 | 5′-GCAGCGGAAAGCAAAACCTACATC-3′ |

### Data analysis

The coding regions of *Brassica oleracea FLC* genes were predicted from the sequence information of mRNA in previously reported *FLCs* from *Arabidopsis* and *Brassica* species and also by identifying the consensus splicing sites at the ends of introns. Amino acids were deduced from the putative exon sequences using “BCM searchlauncher” online service (Smith et al. 1996). Multiple alignment of the deduced amino acid sequences of *FLC* genes from *A. thaliana* and *Brassica* species together with *Arabidopsis MAF* genes were performed using ClustalW (Thompson et al. 1994). Percentage similarities between *FLC* paralogues were calculated using Align (http://www.ebi.ac.uk/emboss/align). Phylogenetic relationships were conducted using MEGA version 3.1 (Kumar et al. 2004) with neighbour-joining based on Kimura’s (1980) two-parameter model. Values were estimated from 500 bootstrap replicates. Highly conserved DNA segments among *FLC* genes were detected by DNA Block Aligner (http://www.ebi.ac.uk/wise2/dbaform.html). Putative gene promoter regions were identified using the PLACE database (Higo et al. 1999), the TSSP (Softberry, http://www.softberry.com) and the TSSP-TCM (Shahmuradov et al. 2005).

## Results

The *B. oleracea* BAC library (JBo) was probed with a fragment (1.2 kb) of *BoFLC3* (AY115673) containing exon 3 to exon 6 and three clones (032J18, 043I02 and 019H24), containing *B. oleracea FLC* genes *BoFLC1*, *BoFLC3* and *BoFLC5*, respectively, were identified by locus-specific PCR. Full-length genomic sequences of these three *BoFLC* genes were obtained and genomic walking subsequently isolated different lengths including 5′-upstream and 3′-downstream of these genes providing approximately 6.2, 5.1 and 8.2 kb of the genomic regions of A12 covering *BoFLC1*, *3* and *5*, respectively. These sequences have been submitted to the EMBL/GenBank databases under the following accession numbers: AM231517 (*BoFLC1*), AM231518 (*BoFLC3*) and AM231519 (*BoFLC5*).

Different primers in different parts of the fully sequenced *B. oleracea FLC4* gene (*BoFLC4-1*: AY306124) were designed (Table 1) to amplify this gene in A12 and GD, but none amplified any fragment in A12. However, we obtained a sequenced fragment (1.2 kb) containing exon 2 to exon 6 of *BoFLC4* (AM231524) from GD. In order to predict the deduced amino acids, exon–intron splicing sites were assumed to be the same as *BoFLC4-1*. A single base pair deletion was detected in exon 4 of the GD allele of *BoFLC4*. It created a frame-shift mutation, which eventually led to an in-frame stop codon (TAA) in exon 4 which

may cause premature termination. Because *BoFLC4* was not identified in A12, we would not expect to detect BAC clones containing *BoFLC4* from a BAC library constructed from A12 genome.

The structures of the three *B. oleracea FLC* paralogues, as well as the existing fully sequenced *BoFLC4-1* (AY306124), were compared with each other and with *Arabidopsis FLC* (Table 2). They all showed the same structure with seven exons. Each particular exon was of similar size among *Arabidopsis FLC* and *BoFLCs*. The genes differed in total size mainly because of the highly variable sizes of introns 1 (fourfold) and 6 (sixfold). *BoFLC5* has an additional nucleotide in exon 2 creating a premature stop codon and it also has an additional nucleotide in exon 7. The *BoFLC3* copy from A12 was slightly larger than *BoFLC3-2* (AY306125), which was previously sequenced in *B. oleracea* var. *capitata* (Lin et al. 2005), mainly due to a large gap (255 bp) in intron 1, but the coding regions were identical.

Overall, high nucleotide sequence conservation was observed within the coding regions among the three *FLC* genes. They showed nucleotide identity between 83 and 88% to *Arabidopsis FLC* in the coding regions (Table 3) but exons 4 and 7 showed lower identities to *Arabidopsis FLC,* ranging 73–82%. The least conservation was between exon 6 of *BoFLC1* and *BoFLC5* (71%). The degree of intron identity could be reliably determined only for introns 3 and 4 because they were conserved in length while the others varied widely in size. Intron 3 identity ranged from 93% between *BoFLC3* and *BoFLC5* to 76% between *Arabidopsis FLC* and *BoFLC1*. *BoFLC3* and *BoFLC5* carried the most conserved intron 4 with 86% identity. Intron 4 from *BoFLC4-1* had less similarity with other *FLC* paralogues, ranging 55–66% because of its size difference.

*Brassica oleracea FLC* copies exhibited high levels of identity to each other (83%) and also to *Arabidopsis FLC* (65–85%) at the amino acid level (Table 4). The degree of similarity was even higher indicating that conserved amino acids have been substituted in some sites. For example, when the predicted amino acids of *Arabidopsis MAF1* were compared to *BoFLC* genes, the identity level significantly decreased whilst amino acid similarity showed a smaller reduction.

## Phylogenetic analysis

Phylogenetic relationships were inferred using the aligned amino acid sequences of several *FLC* genes as well as *MAF* genes. *Arabidopsis MAF* genes were used as an "outgroup" and comprised five genes; *MAF1* (also called *FLM* or *AGL27*) is located on *Arabidopsis* chromosome 1 (At1) and the other four genes (*MAF2-MAF5*) are located in a tight cluster at the bottom of At5 (Ratcliffe et al. 2003). Previous phylogenetic analyses revealed that they constitute the most similar gene family to *FLC* genes whilst being classified as a separate group supported by a high bootstrap value (Tadege et al. 2001; Parenicova et al. 2003). Because the sequences of exon 1 in *B. rapa FLC1*, *FLC2*, *FLC3* and *FLC5* genes were unavailable, a fully amino-acid-sequenced *B. rapa FLC* gene (*BrsFLC*: AAP31678) was added to the alignment to clarify the degree of similarity across the entire gene. The resulting neighbour-joining tree for these genes is shown in Fig. 1.

**Table 2** Intron and exon sizes in *Arabidopsis FLC* (*AtFLC*) (AF116528) and four *B. oleracea FLC* paralogues [*BoFLC1* (AM231517), *BoFLC3* (AM231518), *BoFLC4-1* (AY306124) and *BoFLC5* (AM231519)]

| | Exon 1 | Intron 1 | Exon 2 | Intron 2 | Exon 3 | Intron 3 | Exon 4 | Intron 4 | Exon 5 | Intron 5 | Exon 6 | Intron 6 | Exon 7 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *AtFLC* | 185 | 3493 | 58 | 178 | 62 | 90 | 100 | 78 | 42 | 194 | 42 | 992 | 102 | 4757 |
| *BoFLC1* | 185 | 2342 | 58 | 433 | 62 | 78 | 100 | 85 | 42 | 225 | 42 | 177 | 105 | 3934 |
| *BoFLC3* | 185 | 1364 | 58 | 438 | 62 | 87 | 100 | 80 | 42 | 234 | 42 | 493 | 105 | 3290 |
| *BoFLC4* | 185 | 1123 | 61 | 208 | 62 | 85 | 97 | 65 | 42 | 236 | 42 | 1171 | 105 | 3482 |
| *BoFLC5* | 185 | 4537 | 59 | 410 | 62 | 85 | 97 | 81 | 42 | 77 | 42 | 970 | 106 | 6753 |

AF and AM codes refer to EMBL/GenBank accession numbers

**Table 3** Nucleotide sequence comparisons of *B. oleracea FLC* paralogues and *Arabidopsis thaliana FLC* (% identity) within coding region and also 272 bp 5′ upstream of ATG

| | *AtFLC* | | *BoFLC1* | | *BoFLC3* | | *BoFLC4* | |
|---|---|---|---|---|---|---|---|---|
| | Coding region | 272 bp upstream | Coding region | 272 bp upstream | Coding region | 272 bp upstream | Coding region | 272 bp upstream |
| *BoFLC1* | 88.1 | 58.1 | | | | | | |
| *BoFLC3* | 87.3 | 70.4 | 88.0 | 62.1 | | | | |
| *BoFLC4* | 85.3 | 63.9 | 86.4 | 55.4 | 88.1 | 57.5 | | |
| *BoFLC5* | 82.8 | 62.7 | 83.6 | 46.8 | 86.4 | 67.6 | 83.2 | 54.8 |

**Table 4** Percentage of identities and similarities (in parentheses) of amino acid sequences of *A. thaliana FLC*, *A. thaliana MAF1* and three *B. oleracea FLC* paralogues

| | *AtFLC* | *BoFLC1* | *BoFLC3* | *BoFLC4* |
|---|---|---|---|---|
| *BoFLC1* | 85.4 (91.4) | | | |
| *BoFLC3* | 83.3 (90.9) | 83.2 (90.4) | | |
| *BoFLC4* | 80.9 (91.0) | 82.8 (88.9) | 83.3 (91.9) | |
| *AtMAF1* | 65.3 (80.1) | 66.0 (79.2) | 62.4 (78.4) | 65.7 (79.3) |

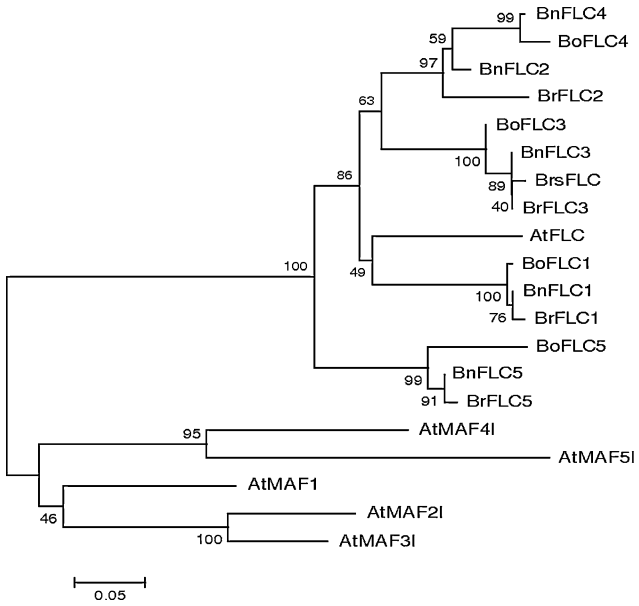


**Fig. 1** Neighbour-joining phylogenetic analysis for the deduced amino acids of *A. thaliana MAF* (*AtMAF1*: AAK37527; *AtMAF2I*: AAO65307; *AtMAF3I*: AAO65310; *AtMAF4I*: AAO65315; *AtMAF5I*: AAO65320), *A. thaliana FLC* (*AtFLC*: AAD21249) and 14 *Brassica FLC* genes (*BoFLC1*: CAJ77613; *BoFLC3*: CAJ77614; *BoFLC4*: AAQ76275; *BoFLC5*: CAJ77618; *BrFLC1*: AAO13159; *BrFLC2*: AAO86066 + AAO86067; *BrFLC3*: AAO13158; *BrFLC5*: AAO13157; *BrsFLC*: AAP31678; *BnFLC1*: AAK70215; *BnFLC2*: AAK70216; *BnFLC3*: AAK70217; *BnFLC4*: AAK70218; *BnFLC5*: AAK70219). *Numbers* at the nodes denote bootstrap support (%) out of 500 replicates

Phylogeny reconstruction showed that *Brassica FLC* genes and *Arabidopsis FLC* grouped together and were well separated from the *MAF* genes, indicating that all *Brassica FLC* copies are orthologues of *Arabidopsis FLC*. Therefore, the neighbour-joining tree represents a monophyletic group including all *FLC* genes with high bootstrap value (100). *Brassica FLC* genes fell within four well-supported clades. Each clade comprised *FLC* genes from diploid *Brassicas* and amphidiploid *Brassica napus*, supporting the consensus view that gene duplication occurred before the divergence of the diploid *Brassica* species. It is consistent with the previous results reported by Lagercrantz and Axelsson (2000) and Rana et al. (2004). *BnFLC4* and *BoFLC4* formed a subgroup within the clade, which also contained *BnFLC2* and *BrFLC2* indicating that *FLC2* and *FLC4* copies are very similar to each other.

The neighbour-joining tree also showed that the three clades containing *BoFLC1*, *BoFLC3* and *BoFLC4* were derived from a common ancestral gene supported by a relatively high bootstrap value (86) but with weak internal resolution. Three of the five identified *BnFLC* genes (*BnFLC1*, *BnFLC3* and *BnFLC5*) are sisters to *B. rapa FLC* copies. *BnFLC4* and *BoFLC4* were sisters to each other supported by a high bootstrap value. The *BnFLC2* relationship remains uncertain due to poor resolution.

## Analysis of promoter regions

The regions, approximately 1.4, 1.8 and 0.7 kb in size carrying the putative promoter elements upstream of the translational start site were sequenced in *BoFLC1*, *BoFLC3* and *BoFLC5*, respectively (Fig. 2). The previously sequenced regions approximately 2.7 and 2.0 kb in size upstream of *Arabidopsis FLC* and *BoFLC4-1* genes were also used to compare the promoters of all four *Brassica FLC* genes with the relatively well characterised *Arabidopsis FLC* promoter. Previous results indicated that 272 bp of the promoter region, upstream of the ATG codon, in combination with the intragenic region, are required to initiate *Arabidopsis FLC* transcription in non-vernalized conditions and also its initial down-regulation induced by cold treatment (Sheldon et al. 2002). The degree of similarity within this critical 5′ upstream segment was evaluated between *B. oleracea FLC* copies and *Arabidopsis FLC* (Table 3). Unlike coding regions, the level of conservation was low and quite variable (47%–70%). Compared to *Arabidopsis FLC*, *BoFLC3* shared a high level of nucleotide identity (70%) and *BoFLC1* showed the least (58.1%). Low similarity within the 272 bp of 5′ upstream sequence was observed even among *BoFLC* copies, although the *Arabidopsis* region and the 5′ regions of *Brassica FLC*s are highly conserved for the first ∼150 bp, which contains transcribed but not translated regions (Fig. 2).

A 75 bp region between −197 and −272 bp upstream of *Arabidopsis FLC* had been reported as essential for non-vernalized expression of *FLC* (Sheldon et al. 2002). High conservation between this region in *Arabidopsis* and the three *BoFLC* copies (*BoFLC3, BoFLC4* and *BoFLC5*) was revealed using "DNA block aligner". *BoFLC3* and *BoFLC5* showed high homology (∼75%) with almost the whole 75 bp region in *Arabidopsis*. *BoFLC4* showed very high homology (85%) with the first 42 bp of the 75 bp region of *Arabidopsis FLC*. While the start positions of conserved regions were within 272 bp upstream of the *BoFLC3* and *BoFLC5* sequences, the conserved region began at position −315 bp in *BoFLC4* but had interruptions to the sequence (Fig. 2).

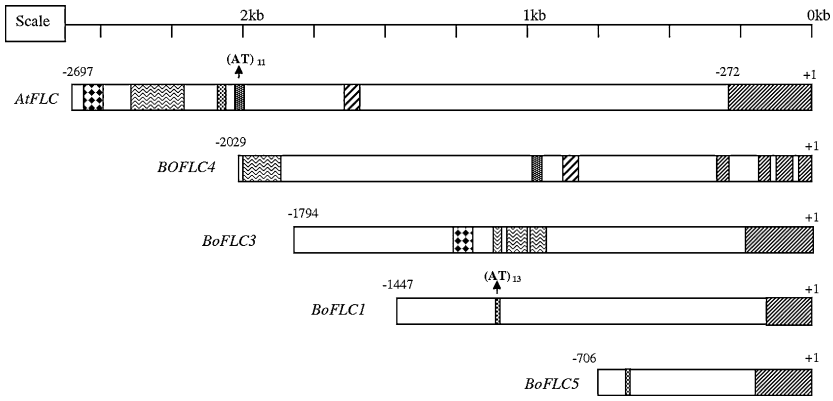


**Fig. 2** Ideogram of the conserved segments (at least 75% identity; shaded) within different regions of 5′ sequence upstream of translational start site (+1) of *A. thaliana FLC* and four *B. oleracea FLC* paralogues

Comparing the promoter sequences of *BoFLC* genes to *Arabidopsis FLC* revealed that the level of similarity decreased in the extended region upstream of the ATG codon and only a few stretches of nucleotides were conserved among *FLC* paralogues (shaded in Fig. 2). We did not detect any conserved segments between the promoter regions of *B. oleracea FLC* genes and those parts of the *FLC* promoter in *Arabidopsis* (−272 to −526 and −678 to −1031), which had been shown previously to have negative and positive regulatory elements (Sheldon et al. 2002). *Arabidopsis FLC* displayed an $(AT)_{11}$ microsatellite sequence 2,135 bp upstream of the ATG codon and an extended version of this, $(AT)_{13}$, was observed 983 bp upstream of the *BoFLC1* start codon (Fig. 2). A stretch of 60 nucleotides was identified in *BoFLC5* showing high conservation to the region of *BoFLC1*, which contains the microsatellite. Lin et al. (2005) identified a gene encoding 3-keto-acyl-ACP dehydratase about 0.9 kb upstream of the translational start site of their *BoFLC3-2* but we did not detect any such gene within 1.8 kb upstream of A12 *BoFLC3*.

The region upstream of the ATG codon of *FLC* genes contains putative promoter motifs, which may be implicated in responses to different environmental factors. In *Arabidopsis FLC*, the nucleotide sequences upstream of the start codon do not contain a TATA box, whilst a TATA-less promoter site was predicted at the beginning of the 5′ untranslated region by TSSP-TCP programme (Shahmuradov et al. 2005). One CAAT box at −247 bp within the essential segment of the promoter region for *FLC* expression and three more CAAT boxes beyond this region at positions −290, −351 and −359 bp were identified. *BoFLC1*, *BoFLC3* and *BoFLC4* had a putative TATA box −181, −188 and −209 bp upstream of the translational initiation signal. *BoFLC5* did not have a TATA box in its promoter region. A TATA-less promoter site was identified at 151 bp upstream of the ATG codon in *BoFLC5*. CAAT boxes were identified in the sequences further upstream of the start codon in *Brassica FLC* genes. *BoFLC1* displayed a CAAT domain at −377 bp. Three CAAT boxes, close to one another, were identified in the promoter region of *BoFLC3* at positions −596, −619 and −636 bp. Also, two CAAT motifs in *BoFLC4* promoter region at −563 and −863 bp and two in *BoFLC5* promoter region at −384 bp and −391 bp were identified.

## Analysis of intron regions

In addition to the 2.0 kb promoter region, the 3.7 kb exon1–intron1–exon2 segment of *Arabidopsis FLC* has been known to contain essential regulatory regions (Sheldon et al. 2002; Bastow et al. 2004). As mentioned earlier, exon regions were very similar among all *FLC* copies. The intron1 sequence of the A12 allele of three *B. oleracea FLC* genes (*BoFLC1*, *BoFLC3* and *BoFLC5*) was compared to intron 1 of *Arabidopsis FLC* which had already been characterized in terms of regulating various *FLC* activities (Sheldon et al. 2002). Several segments within intron 1 of *Arabidopsis* and *Brassica FLC* genes could be aligned with a high level of identity. Figure 3 shows the conserved segments between intron 1 of *Brassica FLC* genes and *Arabidopsis FLC* with at least 75% identity. As reported for *BoFLC4-1* (Lin et al. 2005), the three *Brassica FLC* genes have conserved segments with fragments A and F of intron 1 of *Arabidopsis FLC* which are presumably involved in two major *FLC* activities, non-vernalized expression and its repression by vernalization. It is known that some parts of the 2.8 kb region of *Arabidopsis FLC* intron 1 containing segments B, C, D and E are required for the maintenance of *FLC* repression induced by prolonged low temperature (Sheldon et al. 2002). The segments which show high conservation with fragments B and C of *Arabidopsis FLC* intron 1 were found within intron 1 of *B. oleracea FLC* copies. No conserved region was identified between segments E of *Arabidopsis FLC* intron 1 and those *B. oleracea FLC* genes, which have a shorter intron 1 sequence than that in *Arabidopsis*. Within the *BoFLC5* intron 1, large segments could be observed which did not contain any conserved segments with the other *FLC* genes.

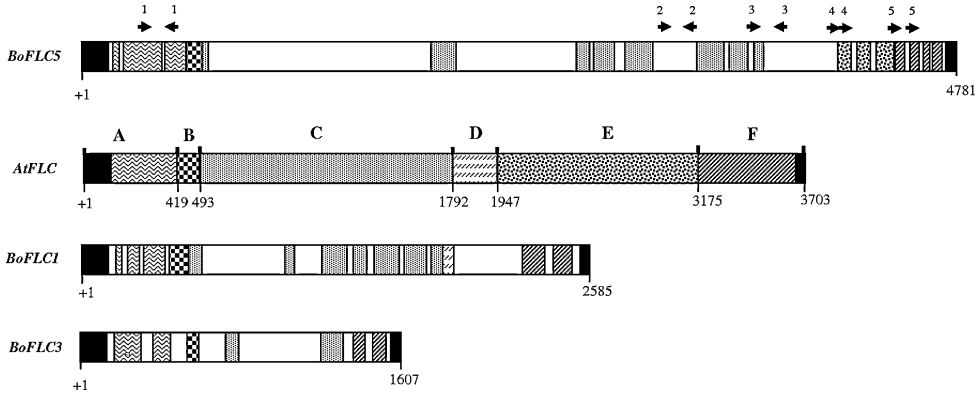


**Fig. 3** Ideogram of intron 1 of *A. thaliana FLC* and *B. oleracea FLC* paralogues showing conserved segments. Bordering exon regions are shown in *black*. Intron 1 of *AtFLC* is partitioned to show regulatory segments, *shaded by different patterns*, involved in various *FLC* activities (Sheldon et al. 2002). Segments A and F are essential for non-vernalized expression and FLC down-regulation by vernalization. The other segments are involved in maintenance of *FLC* low expression. Direct sequence repeats within the *BoFLC5* intron 1 are shown by *arrows* with the same direction and the three inverted sequence repeats are shown by *arrows* with opposite direction

Two tandem direct repeats, 40 and 39 bp in size, starting at positions (+4069, +4109) and (+4525, +4567) are present along the *BoFLC5* intron1. Three inverted repeats; 14, 35 and 16 bp in size; were also identified in intron 1 of *BoFLC5* located at (+274, +486), (+3099, +3328) and (+3540, +3788). The positions are relative to ATG as +1. No duplications were detected in other *FLC* genes.

## Evaluating *BoFLC* genes as flowering time QTL candidate genes

### *Allelic differences between the parental lines*

In addition to the early flowering parental line (A12), the sequences of all predicted exon regions of *BoFLC1*, *BoFLC3* and *BoFLC5* were obtained in the late flowering parental line (GD) in order to assess whether *FLC* genes could be involved in causing flowering time variation in this population of *B. oleracea*. Two fragments of the GD allele of each *BoFLC* gene were sequenced; one covered exon 1 (>500 bp of each *BoFLC*) and the other contained exon 2 to exon 7 (1.7 kb for *BoFLC1*, 2.0 kb for *BoFLC3* and 2.1 kb for *BoFLC5*). Primers were designed from the A12 gene sequences. All six sequenced fragments were placed in EMBL/GenBank databases (AM231520-AM231523, AM231525 and AM231526).

Comparisons between the A12 and GD alleles of *BoFLC1* and *BoFLC3* revealed a number of amino acid differences (Fig. 4). Four nucleotide changes were observed within *BoFLC1* coding regions between A12 and GD allele. One was a synonymous substitution but three were non-synonymous substitutions, which may confer different physico-chemical properties to the resulting protein. They are residue 63 (Leu to Pro), residue 85 (Ser to Leu), and residue 176 (Tyr to Asn) Within the *BoFLC3* coding sequences, GD showed four nucleotide differences in comparison with A12; three resulted in amino acid polymorphism, whereas one, in exon 7, was a synonymous substitution. A non-conserved amino acid substitution (Gly-110 to Val) was identified in the K-box, which is a coiled-coil region of the *FLC* protein. Another two amino acid differences, Lys-74 to Asn and Lys-151 to Arg, were conserved substitutions, which may be functionally silent. Only one nucleotide difference was detected within *BoFLC5* coding regions between A12 and GD, which made no change at the amino acid level. As in A12, the GD allele of *BoFLC5* contained an in-frame stop codon in exon 2.

### *Locating paralogous* FLC *loci using the selective set of DH lines*

We chose eight DH lines which represented all possible combinations of both parental alleles of the three FT-QTL on linkage groups O2, O3 and O9 of *B. oleracea* (Bohuon et al. 1998) based on their estimated 95% confidence intervals. These lines exhibited unique QTL allelic patterns for the three linkage groups of interest (Fig. 5). They were targeted to locate *BoFLC* copies because the QTL confidence intervals on O2 (partially) or on O3 and O9 (entirely) overlapped with the segments collinear with the top of At5 containing *FLC*. The three target fragments provided by the eight selected DH lines were as follows, cM position (marker name):

(a) 56.3 (pR86J1)–106.6 (AC-CTAE03) cM on O2,
(b) 0 (pW116J1)–29.2(pN180E3) cM on O3 and
(c) 67.7(pO160E1)–113 (AC-CATR14) cM on O9.

Figure 5 shows the QTL regions and the identified *FLC* alleles in all eight DH lines. Because the patterns for *BoFLC1* and *BoFLC3* exactly corresponded with QTL alle-


**Fig. 4** Alignment of the deduced amino acid sequences of A12 and GD alleles of two *B. oleracea FLC* paralogues (*BoFLC1* and *BoFLC3*) to show polymorphic sites. *Black bars* indicate non-synonymous differences

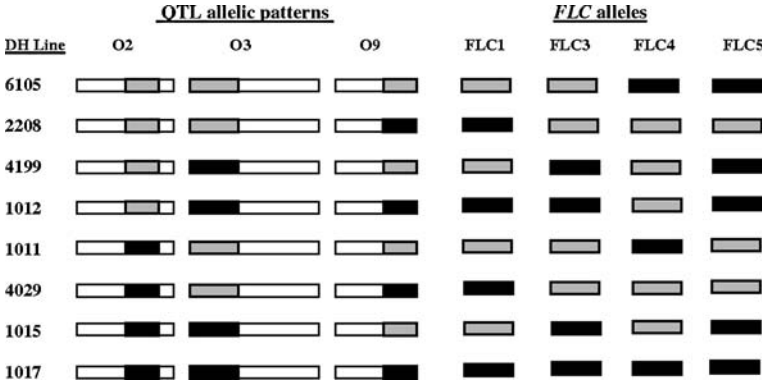


**Fig. 5** **a** QTL allelic patterns in eight DH lines which represent all possible combinations of parental alleles of the segments containing FT-QTL located on linkage groups O2, O3 and O9. **b** The alleles of *FLC1*, *FLC3*, *FLC4* and *FLC5* identified in this orthogonal set of DH lines. *Grey* and *black boxes* correspond to A12 and GD alleles, respectively

lic patterns on O9 and O3, respectively, they were positioned within the target regions (c) and (b), respectively. The pattern for *BoFLC5* was the same as that for *BoFLC3* for seven of the eight lines suggesting that *BoFLC5* is probably also located on linkage group O3. However, the fact that DH6105 displayed a GD allele for *BoFLC5*, even though it had an A12 haplotype for all three target QTL regions, indicates that, in spite of locating to linkage group O3, *BoFLC5* is outside the putative confidence interval of the FT-QTL. *BoFLC5* was located using haplotype data from DH lines 4199 and 6105 between 24.2 and 55.3 cM on linkage group O3. It falls within the confidence interval of a second FT-QTL on O3 situated between 36.5 and 77.3 cM, identified by Rae et al. (1999) using substitution lines.

Using *BoFLC4* primers F1 and R1, a PCR product was obtained from three lines only (DH 1011, DH 1017, and DH 6105), and the sequence from each corresponded to the GD fragment. PCR amplification with other *BoFLC4* primer pairs supported this result. It was assumed that those lines in which *BoFLC4* was not amplified had the A12 genotype. The *BoFLC4* allelic pattern was compared to the QTL allelic pattern for these lines (Fig. 5) but it did not correspond to any of them and, therefore, *BoFLC4* must fall outside these FT-QTL confidence intervals. To locate *BoFLC4*, the haplotype data for these eight lines were searched for any loci which shared an allelic pattern with *BoFLC4* in these eight lines. This occurred at just one locus, pN121E1 at 31.2 cM on linkage group O2. Therefore, the most likely position for *BoFLC4* is on linkage group O2 between markers pW116E1 and pN102E2, 0–38.7 cM.

To locate each *BoFLC* copy more precisely, AG–DH lines with crossovers at different points within these QTL regions were selected and the allelic sequence of the appropriate *BoFLC* gene determined (Fig. 6). A *BoFLC1* fragment was amplified and then sequenced in DH lines (DH2190, DH5079 and DH2056). All three lines had the A12 allele and therefore *BoFLC1* must be situated between 84.6 and 113 cM on O9 (Fig. 6). This agrees with previous results, which located *BoFLC1* on O9 between 87.2 cM (pN47E4NM) and 103.5 cM (pN3E1) (Salathia 2003). The *BoFLC3* fragment was amplified and sequenced in DH lines (DH2056, DH3123, DH4030 and DH5008) and was

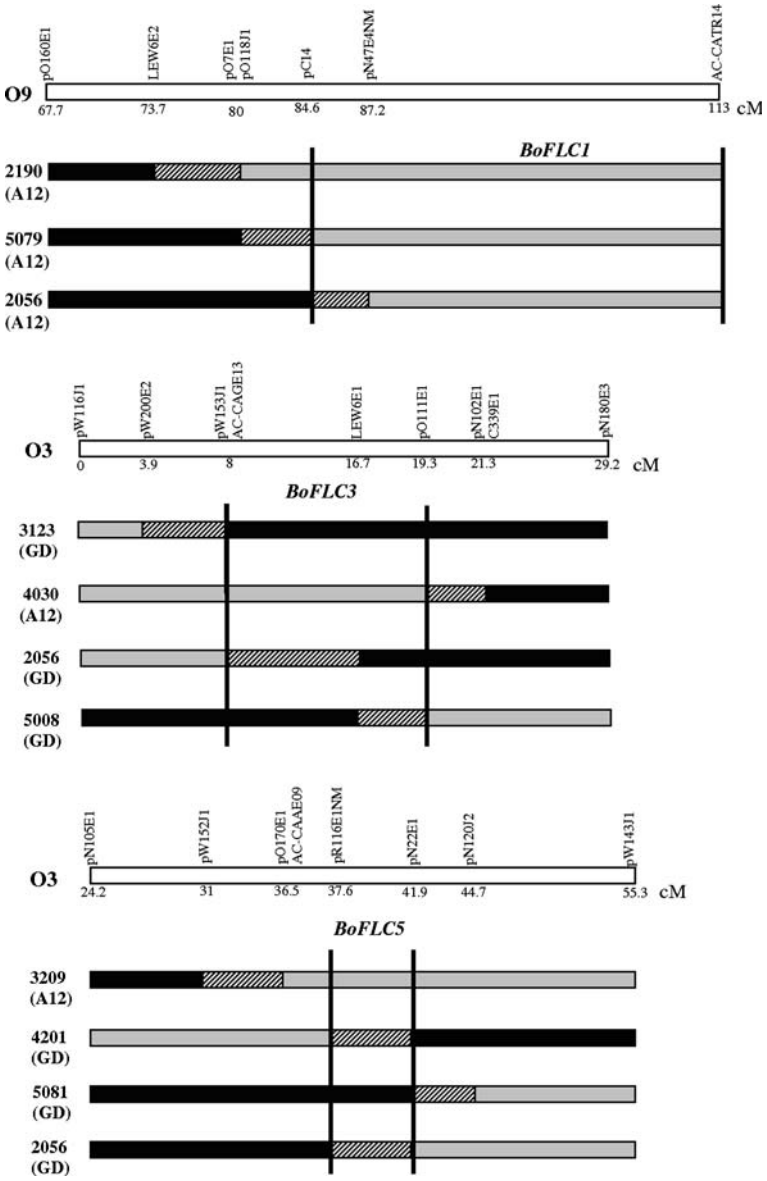


**Fig. 6** Fine mapping of *BoFLC* genes using additional DH lines. GD segments are in *black* and A12 segments are in *grey*. *Hatched areas* show unknown segments. *FLC* genes are located between the *vertical bars*

located between 8 and 19.3 cM on O3 (Fig. 6). Four DH lines (DH2056, DH3209, DH4201 and DH5081) were used to locate *BoFLC5* between 37.6 and 41.9 cM on O3 (Fig. 6). Because there is no marker mapped between 0 and 31.2 cM on linkage group O2, *BoFLC4* cannot be located more precisely at present. The locations of these four *FLC* genes in relation to FT-QTL are summarised in Fig. 7. The use of a common set of RFLP markers mapped onto *Arabidopsis* chromosomes (Parkin et al. 2005) and the integrated genetic map from *B. oleracea* (Sebastian et al. 2000) allowed us to define the collinear segments between the top of At5 and *B. oleracea* linkage groups O2, O3 and O9. The tracts (and corresponding flanking markers) homologous with the top of At5 shown in Fig. 7 are as follows: 0.0 (pW116J1)–56.3 (pR86J1) cM on O2, 0.0 (pW116J1)–31.0 (pW152J1) cM on O3 and 60.8 (pN180E1)–103.5 (pW200J1) cM on O9. Three *FLC* copies in *B. oleracea* (*BoFLC1*, *BoFLC3* and *BoFLC4*) were situated within the predicted regions collinear with the top of At5 suggesting that they may have arisen by genome polyploidisation. In contrast, the *BoFLC5* location (O3; 37.6–41.9 cM) was outside these collinear segments.

### FLC *segregations and FT in backcross progeny from substitution lines*

To pursue the role of *B. oleracea FLC* genes as possible candidates for FT-QTL variation further, the allelic source (A12 vs. GD) of three *BoFLC* genes were explored among the three selected substitution lines and the late flowering individuals of generation $BC_1S_1$ derived from these lines. The presence of individuals among $BC_1S_1$ families, which

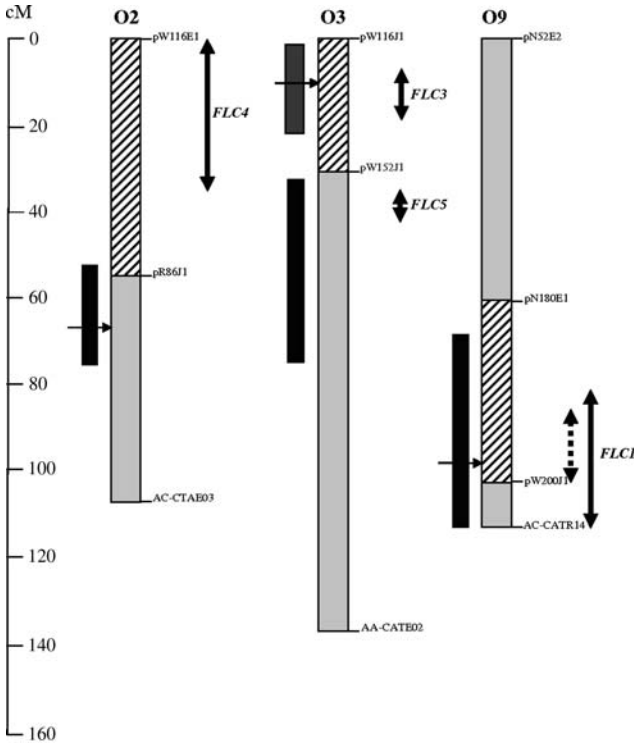


**Fig. 7** The position of four *FLC* loci on linkage groups O2, O3, and O9 in *B.oleracea* in relation to FT-QTL locations. *Hatched areas* show syntenic regions with the top of At5. *Black bars* correspond to FT-QTL confidence intervals with horizontal arrows indicating most likely position. The *dotted arrow* for *FLC1* is the location given by Salathia (2003)

differed significantly in flowering time from A12, implies that the GD fragment of each substitution line contains QTL, which affect flowering time.

We genotyped 15 $BC_1S_1$ individuals that were significantly later flowering than A12, to check if the homozygotes were preponderantly GD for the *BoFLC* genes. SL128, which had a single GD fragment (8–43.3 cM) on linkage group O3, carried the A12 allele of *BoFLC3* (8–19.3 cM) strongly suggesting that a gene and or genes other than *BoFLC3* caused late flowering in this line. As expected, therefore, all the late flowering individuals among the BC1S1 also had the A12 allele of *BoFLC3*. Both SL133 and SL175 carried the GD allele for *BoFLC5* and *BoFLC1*, respectively, implying that they could be considered as potential candidate genes for late flowering. However, both A12 and GD alleles were approximately equally represented among the late flowering progeny from *BoFLC1* and *BoFLC5*, suggesting that *FLC* is not responsible for their late flowering.

## Discussion

The presence of replicated copies of most genes is a general characteristic of diploid *Brassica* species and may contribute to novel phenotypic variation within them (Lukens et al. 2004). In this study, *B. oleracea FLC* paralogues were isolated and sequenced in a rapid cycling line (A12) and a late flowering line (GD) to discover if *BoFLCs* show variation that might support their being candidates for FT-QTL. Such sequence information may also lead us to a better understanding of the entire gene structure and their promoter regions with respect to *Arabidopsis FLC*. It can be argued that *FLC* does not constitute a strong candidate gene for FT-QTL in the context of the AG population because they are annuals and were grown in the absence of cold stress. However, given that the *Brassica* QTL regions syntenous to the top of At5 have been found in a wide range of species including those that do and that do not require cold treatment to flower, it is important to test whether or not they are candidates here. They also offered the opportunity for a thorough sequence comparison across four ancient paralogues.

Prior to this work, four partially sequenced *B. oleracea FLC* copies [*BoFLC1* (AY115672), *BoFLC2* (DQ222849-DQ222850), *BoFLC3* (AY115673) and *BoFLC5* (AY115674) and two fully sequenced *B. oleracea FLC* copies {*BoFLC3-2* (AY306123) and *BoFLC4-1* (AY306122)] had been placed in the GenBank/EMBL databases

DNA sequence comparisons showed similar structure and high homology in the coding regions both between the paralogous *BoFLC* genes and between them and *Arabidopsis FLC*. Although the A12 alleles of *BoFLC3* (AM231518) and *BoFLC3-2* (AY306123) represented identical coding sequences, they differed in intron and promoter regions. This suggests that unlike *BoFLC3-2*, A12 *BoFLC3* may be expressed because its promoter region (1.8 kb) showed two large, unique DNA inserts, which may contain necessary regulatory elements, not present in *BoFLC3-2* (Lin et al. 2005). Moreover, the promoter region of the A12 allele of *BoFLC3* was not interrupted by the sequence of another gene, as occurred in *BoFLC3-2* (Lin et al. 2005). It is possible, therefore, that two variants of *FLC3* may exist within the *B. oleracea* genome; two tandemly located *FLC3* have been found in *B. rapa* (Yang et al. 2006). Similarly, two copies of *BoFLC1* have been reported within a single BAC clone (JBo032J18) (Salathia 2003).

Phylogenetic analysis confirmed that *Brassica FLC* copies are *Arabidopsis FLC* orthologues as previously reported by Schranz et al. (2002), Martynov and Khavkin (2004) and Lin et al. (2005). *MAF* genes form a separate group, well differentiated from the *FLC* clade. The functional divergence observed between *Arabidopsis FLC* and *MAF* genes might be explained by differences in their amino acid sequences. It is interesting to note that amino acid residue 30, which is occupied by two polar amino acids (Asp or Glu) in both *Arabidopsis FLC* and *MAF* genes, was also conserved in *Brassica FLC* and they all had Glu at this

position. Conversely, all other *Arabidopsis* genes encoding MADS domain proteins other than *FLC* and *MAF* genes produce a positively charged amino acid (Lys) at residue 30. These features suggest that residue 30 could confer specific DNA binding properties to *Arabidopsis FLC* and *MAF* proteins (Ratcliffe et al. 2001).

Phylogenetic reconstruction supports the view that the *FLC* gene duplications occurred before *B. oleracea* and *B. rapa* diverged about 4 MYA (Rana et al. 2004) because each clade contained *FLC* copies from different *Brassica* species. This agrees with the results of Schranz et al. (2002), Lin et al. (2005) and Okazaki et al. (2007). Furthermore, there is some suggestion that gene duplication events had taken place before the divergence of *Brassica* and *Arabidopsis* lineages because *FLC* genes, including *Arabidopsis FLC*, formed a monophyletic group. Previous work showed no consensus outcome for *Arabidopsis FLC* placement within the phylogenetic tree (Schranz et al. 2002; Martynov and Khavkin 2004; Li et al. 2005; Lin et al. 2005; Okazaki et al. 2007). This may be because they used different lengths of aligned coding sequences or amino acid sequences to reconstruct the phylogenetic tree. Therefore, the different numbers of polymorphic sites recognized by the tree-building methods might cause *Arabidopsis FLC* to locate to different positions in the trees.

Comparisons between promoter and intron regions of *Arabidopsis FLC* and *B. oleracea FLC* copies demonstrated that those parts which are essential for normal *FLC* expression and vernalization-induced *FLC* repression in *Arabidopsis* were conserved in all *BoFLC* genes, indicating that similar sequence motifs are involved in controlling major *FLC* activities in both species. Whether there are any *cis*-acting elements, which are specifically involved, in down-regulating *FLC* expression in response to vernalization remains to be investigated. On the other hand, lack of conservation of the segments found in the *Arabidopsis FLC* promoter region, which are known to contain positive and negative regulatory elements (−272 to −526 bp and −687 to −1031 bp upstream of the ATG codon) (Sheldon et al. 2002), may suggest different regulatory mechanisms in *B. oleracea* and *Arabidopsis FLC* genes, probably in connection with the distinct vernalization response between *B. oleracea* (plant-vernalization-responsive) and *Arabidopsis* (seed-vernalization-responsive), as previously described by Lin et al. (2005).

Whether *BoFLC* genes are the candidates for their respective flowering time QTL was a key question of this project. The presence of QTL affecting flowering time in the regions of a *B. oleracea* genetic map collinear with the top of *Arabidopsis* chromosome 5 containing *FLC* (Bohuon et al. 1998; Rae et al. 1999) was the reason underlying our exploration of *BoFLC* paralogues as putative candidate genes. From this point of view, two principal questions were raised; first whether *BoFLC* genes are located within the confidence intervals of their corresponding QTL and second, whether allelic variation of *BoFLC* loci exists between parental lines (A12 and GD) of the mapping population which has led to the identification of the flowering time QTL. Our mapping of four *BoFLC* loci onto the genetic map confirmed that three *BoFLC* copies (*BoFLC1*, *BoFLC3* and *BoFLC5*) are located within the confidence interval of their respective FT-QTL (Bohuon et al. 1998; Rae et al. 1999), while *BoFLC4* is located on a region of linkage group O2 where no flowering time QTL was detected. Conversely, Okazaki et al. (2007) detected an FT-QTL with large effect in the region of linkage group O2 where their *BoFLC2*, which is probably the same as *BoFLC4*, was mapped. It should be noted that Okazaki et al. (2007) detected an allele of *BoFLC2* with no premature stop codon in one of the parental lines, while our mapping population was generated from two lines, one (A12) with no *BoFLC4* allele and another line (GD) with a *BoFLC4* allele containing a premature stop codon in exon 4 resulting from a single base deletion. This deletion was also identified in some cultivars of broccoli and cauliflower (Okazaki et al. 2007).

*BoFLC5* (AM231519) almost certainly represents a pseudogene due to a premature in-frame stop codon in exon 2. It also contains a very large intron 1 (4,537 bp) which is almost as large as intron 1 of the *FLC* allele in *Arabidopsis* accession *Landsberg erecta* (*Ler*) (4,695 bp). Lack of *BoFLC5* expression had also been reported in a rapid cycling *B. oleracea*, which did not have any pre-existing stop codon within exon 2–7 (Pires et al. 2004; Okazaki et al. 2007). Lack of polymorphism in coding regions between *BoFLC5* alleles of an early (A12) and a late flowering line (GD) implied that it was not responsible for the differences in flowering time. So, other genes must underlie the FT-QTL located in the *FLC* region of O3.

Both *BoFLC1* and *BoFLC3* were polymorphic between A12 and GD parents for non-synonymous amino acid codons that could affect FT. Okazaki et al. (2007) did not detect any link between *BoFLC1* and *BoFLC3* genes and FT-QTL in their population. The lack of any correlation between the segregation of GD alleles of *BoFLC1* and *BoFLC3* with late flowering in the $BC_1S_1$ progenies strongly argues against the genes being FT-QTL candidates in our population. Moreover, both *BoFLC5* and *BoFLC4* appear to be pseudogenes. Thus, we have strong evidence against any identified *BoFLC* being candidates for the FT-QTL in our population. Of course, this does not preclude polymorphism in *FLC* paralogues being responsible for FT-QTL in other populations of *B. oleracea* or, indeed, in other species.

**Acknowledgments** This work was supported by a grant from the Iranian Ministry of Science, Research and Technology to H. Razi, We would like to express our thanks to colleagues at Warwick HRI (G. Barker, G. King, G. Teakle) for access to the BACs and general advice.

## References

Axelsson T, Shavorskaya O, Lagercrantz U (2001) Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. Genome 44:856–864

Ayele M, Haas BJ, Kumar N, Wu H, Aken SV, Utterback TR, Wortman JR, White OR, Town CD (2005) Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in *Arabidopsis*. Genome Res 15:487–495

Bastow R, Myine JS, Lister C, Lippman Z, Martienssen RA, Dean C (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427:164–167

Bohuon EJR, Ramsay LD, Craft JA, Arthur AE, Marshal DF, Lydiate DJ, Kearsey MJ (1998) The association of flowering time quantitative trait loci with duplicated regions and candidate loci in *Brassica oleracea*. Genetics 150:393–401

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome 41:62–69

Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiol 132:1107–1114

Henderson IR, Shindo C, Dean C (2003) The need for winter in the switch to flowering. Annu Rev Genet 37:371–392

Hepworth SR, Valverde F, Ravenscroft D, Mouraduv A, Coupland G (2002) Antagonistic regulation of flowering time gene SOC1 by CONSTANS and FLC via separate promoter motifs. EMBO J 21:4327–4337

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297–300

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kole C, Quijada P, Michaels SD, Amasino RM, Osborn TC (2001) Evidence for homology of flowering time genes VFR2 from *Brassica rapa* and FLC from *Arabidopsis thaliana*. Theor Appl Genet 102:425–430

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228

Lagercrantz U, Axelsson T (2000) Rapid evolution of the family of *CONSTANS LIKE* genes in plants. Mol Biol Evol 17:1499–1507

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in *Brassica*. Genetics 144:1903–1910

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9:13–20

Levy YY, Dean C (1998) The transition to flowering. Plant Cell 10:1973–1989

Li Z, Zhao L, Cui C, Kai G, Zhang L, Sun X, Tang K (2005) Molecular cloning and characterization of an anti-bolting related gene (*Brp-FLC*) from *Brassica rapa* ssp. *pekinensis*. Plant Sci 168:407–413

Lin S, Wang J, Poon S, Su C, Wang S, Chiou T (2005) Differential regulation of *FLOWERING LOCUS C* expression by vernalization in cabbage and *Arabidopsis*. Plant Physiol 137:1037–1048

Lukens LN, Quijada PA, Udall J, Pires JC, Schranz ME, Osborn TC (2004) Genome redundancy and plasticity within ancient and recent *Brassica* crop species. Biol J Linn Soc 82:665–674

Martynov VV, Khavkin EE (2004) Two homologs of the *FLOWERING LOCUS C* gene from leaf mustard (*Brassica juncea*). Russ J Plant Physiol 51:234–240

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

Michaels SD, Amasino RM (2000) Memories of winter: vernalization and the competence to flower. Plant Cell Environ 23:1145–1153

Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of Flowering Locus C activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. Proc Natl Acad Sci USA 100:10102–10107

Michaels SD, Himelblau E, Kim SY, Shomburg FM, Amasino RM (2005) Integration of flowering signals in winter-annual *Arabidopsis*. Plant Physiol 137:149–156

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

O'Neil MC, Bancroft I (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoelogous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. Plant J 23:233–243

Osborn TC, Kole C, Parkin IAP, 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

Parenicova L, Folter SD, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. Plant Cell 15:1538–1551

Parkin AP, Lydiate DJ, Trick M (2002) Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. Genome 45:356–366

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781

Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangments in resynthesized *Brassica* polyploids (Brassicaceae). Biol J Linn Soc 82:675–688

Quiros CF, Grellet F, Sadowski J, Suzuki T, Li G, Wroblewski T (2001) *Arabidopsis* and *Brassica* comparative genomics: sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. Genetics 157:1321–1330

Rae AM, Howell EC, Kearsey MJ (1999) More QTL for flowering time revealed by substitution lines in *Brassica oleracea*. Heredity 83:586–596

Ramsay LD, Jenings DE, Bohuon EJR, Arthur AE, Lydiate DJ, Kearsey MJ, Marshall DF (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome 39:558–567

Rana D, Boogaart TV, O'Neill CM, Hynes L, Bent E, Macpherson L, Park JY, Lim YP, Bancroft I (2004) Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. Plant J 40:725–733

Ratcliffe OJ, Nadzan GC, Reuber TL, Riechmann JL (2001) Regulation of flowering in *Arabidopsis* by an FLC homologue. Plant Physiol 126:122–132

Ratcliffe OJ, Kumimoto RW, Wong BJ, Riechmann JL (2003) Analysis of the *Arabidopsis MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. Plant Cell 15:1159–1169

Razi H (2006) Characterising paralogues of *FLOWERING LOCUS C* in *Brassica oleracea*. PhD Dissertation, University of Birmingham

Roberts LS, Robson F, Sharpe A, Lydiate D, Coupland G (1998) Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*. Plant Mol Biol 37:763–772

Salathia NS (2003) Regulation of biological clocks in *Brassica oleracea* and *Arabidopsis thaliana*. PhD Dissertation, University of Warwick

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

Sebastian RL, Howell EC, King GJ, Marshal DF, Kearsey MJ (2000) An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. Theor Appl Genet 100:75–81

Shahmuradov IA, Solovyev VV, Gammerman AJ (2005) Plant promoter prediction with confidence estimation. Nucleic Acids Res 33:1069–1076

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, Finnegan EJ, Rouse DT, Tadege M, Bagnall DJ, Helliwell CA, Peacock WJ, Dennis ES (2000a) The control of flowering by vernalization. Curr Opin Plant Biol 3:418–422

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

Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. Plant Cell 14:2527–2537

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C (2005) Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. Plant Physiol 138:1163–1173

Smith RF, Wiese BA, Wojzynski MK, Davison DB, Worley KC (1996) BCM search launcher—An integrated interface to molecular biology data base search and analysis services available on the World Wide Web. Genome Res 6:454–462

Sung S, Amasino RM (2005) Remembering winter: toward a molecular understanding of vernalization. Annu Rev Plant Biol 56:491–508

Suzuki T, Grellet F, Potter D, Li G, Quiros CF (2003) Structure, sequence, and phylogeny of the members of the Ck1 gene family in *Brassica oleracea* and *Arabidopsis thaliana* (Brassicaceae). Plant Sci 164:735–742

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

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of the progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Yang TJ, Kim JS, Lim KB, Kwon SJ, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Kim SH, Lim YP, Park BS (2005) The Korea *Brassica* genome project: a glimpse of the *Brassica* genome based on comparative genome analysis with *Arabidopsis*. Comp Funct Genomics 6:138–146

Yang TJ, Kim JS, Kwon SJ, Lim KB, Choi BS, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Lim YP, Kang JJ, Hong JH, Kim CB, Bhak J, Bancroft I, Park BS (2006) Sequence-level analysis of the diploidization process in the triplicated *FLOWERING LOCUS C* region of *Brassica rapa*. Plant Cell 18:1339–1347