Natural variation in *CIRCADIAN CLOCK ASSOCIATED 1* is associated with flowering time in *Brassica rapa*

Journal:	<i>Genome</i>
Manuscript ID	gen-2016-0052.R2
Manuscript Type:	Article
Date Submitted by the Author:	06-Oct-2016
Complete List of Authors:	Yi, Hankuil; Chungnam National University Li, Xiaonan; Chungnam National University; Shenyang Agricultural University Lee, Seong Ho; Chungnam National University, Department of Horticulture Nou, Ill-Sup; Sunchon National University, Horticulture Lim, Yong Pyo; Department of horticulture, College of agriculture and life sciences, Hur, Yoonkang; Chungnam National University, South Korea,
Keyword:	<i>Brassica rapa</i>, <i>CCA1</i>, Flowering time, PCR markers, Sequence variation

SCHOLARONE™
Manuscripts

Natural variation in *CIRCADIAN CLOCK ASSOCIATED 1* is associated with flowering time in *Brassica rapa*

Hankuil Yi¹⁺, Xiaonan Li^{2,4+}, Seong Ho Lee², Ill-Sup Nou³, Yong Pyo Lim^{2*}, and Yoonkang Hur^{1*}

¹Department of Biological Science, College of Bioscience and Biotechnology, Chungnam National University, Daejeon 305-764, Republic of Korea

²Department of Horticulture, Chungnam National University, Gung-Dong, Yuseong-Gu, Daejeon 305-764, Republic of Korea

³Department of Horticulture, Sunchon National University, Suncheon, Jeonnam, Republic of Korea

⁴Department of Horticulture, Shenyang Agricultural University, Shenyang, China

⁺ Contributed equally to this work.

^{*} Co-corresponding authors.

Yong Pyo Lim

Department of Horticulture

Chungnam National University, Daehak-ro 99, Yuseong-Gu, Daejeon 305-764, Republic of Korea

Tel: 82-42-821-5739

yplim@cnu.ac.kr

Yoonkang Hur

Department of Biological Science

Chungnam National University, Daehak-ro 99, Yuseong-Gu, Daejeon 305-764, Republic of Korea

Tel: 82-42-821-6279

ykhur@cnu.ac.kr

Abstract

Flowering time is a very important agronomic trait and the development of molecular markers associated with this trait can facilitate crop breeding. *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*, a core oscillator component of circadian rhythms that affect metabolic pathways in plants, has been implicated in flowering time control in *Brassica* species. *CCA1* gene sequences from three *Brassica rapa* inbred lines, showing either early flowering or late flowering phenotypes, were analyzed and a high level of sequence variation was identified, especially within the fourth intron. Using this information, three PCR primer sets were designed and tested using various inbred lines of *B. rapa*. The usage of InDel marker was further validated by evaluation of flowering time and high resolution melting (HRM) analysis. Both methods, PCR and HRM, validated the use of newly developed marker. Additional sequence analyses of *Brassica* plants with diploid (AA, BB or CC) and allotetraploid genomes further confirmed a large number of sequence polymorphisms in the *CCA1* gene, including insertions/deletions in the fourth intron. Our results demonstrated that sequence variations in *CCA1* can be used to develop valuable trait-related molecular markers for *Brassica* crop breeding.

Key words: *Brassica rapa*; *CCA1*; Flowering time; PCR markers; Sequence variation.

Author contribution statement

HY, XL, YPL, and YH conceived the design of this study. HY, XL, and ISN carried out experiments and analyses. YPL and YH wrote the manuscript.

Introduction

Flowering in plants is controlled by various factors such as photoperiod, temperature, phytohormones, and the developmental stage of plants, with photoperiod and temperature playing the major roles (Srikanth and Schmid 2011). The involvement of these factors at the molecular level is best understood in *Arabidopsis* (*Arabidopsis thaliana*), in which a number of genes and miRNAs involved in the regulation of flowering time have been identified. In vegetable crops like Chinese cabbage (*Brassica rapa*), as well as in cereals, flowering time is associated with productivity and harvesting time (Cockram et al. 2007; Jung and Müller 2009). In particular, early flowering (EF) in leafy vegetable Chinese cabbages (*B. rapa* ssp. *pekinensis* and *chinensis*) causes a reduction in productivity and market value. Therefore much effort has been made in breeding programs to develop late flowering (LF) Chinese cabbage varieties.

The circadian clock regulates flowering time in concert with the photoperiodic flowering pathway (Jung and Müller 2009; Imaizumi 2010; Song et al. 2013 and 2015). As a core component of the clock, CIRCADIAN CLOCK ASSOCIATED 1 (CCA1), a MYB-related transcription factor, plays an important role in the phytochrome-dependent induction of photosynthetic genes (Wang et al. 1997, Green and Tobin 2002), regulation of the circadian clock (Green and Tobin 2002; McClung 2014), regulation of stress response (Dong et al. 2011; Lai et al. 2012; Seo et al. 2012), and regulation of photoperiodic flowering (Niwa et al. 2007; Fujiwara et al. 2008). LATE ELONGATED HYPOCOTYL (LHY) is another core component of the circadian clock in *Arabidopsis* (Wang et al. 1997; Wang and Tobin 1998) and is involved in the regulation of photoperiodic flowering (Fujiwara et al. 2008; Imaizumi 2010; Li et al. 2011; Lu et al. 2012). In *Capsella bursa-pastoris*, a close relative of *Arabidopsis*, *CCA1* shows different expression patterns in two ecotypes with contrasting flowering phenotypes, suggesting its association with adaptive flowering behavior (Slotte et al. 2007). In a collection of *B. rapa* accessions, most clock quantitative trait loci (QTLs) overlap with flowering time QTLs (Lou et al. 2011). Taken together, these observations suggest that *CCA1* may be involved in the regulation of flowering time in *B. rapa*.

Brassica is one of the most important genera within the Brassicaceae (Cruciferae or mustard family) and includes 38 species and numerous varieties that are grown for agriculture, horticulture, and human

health. Within the *Brassica* genus, six crop species found in the “Triangle of U” are cultivated globally: three basic diploid species, *B. rapa* (A genome, $n = 10$), *B. oleracea* (C genome, $n = 9$), and *B. nigra* (B genome, $n = 8$), which between them gave rise to three amphidiploid species, *B. napus* (AC genome, $n = 19$), *B. juncea* (AB genome, $n = 18$), and *B. carinata* (BC genome, $n = 17$) (U 1935). Of these six *Brassica* species, cultivars of *B. rapa* are grown for leafy vegetables, vegetable oils, turnip roots, turnip greens, turnip tops, and fodder turnip. A number of major cultivar groups were distinguished in the past as subspecies based on their morphological appearance and the plant parts consumed as *B. rapa* ssp. *pekinensis* (heading type), *chinensis* (non-heading type), *parachinensis*, *broccoletto*, *narinosa*, *nipposinica*, *tricoloris*, *oleifera*, and *rapa* (Zhao et al. 2005). Kenshin and Chiifu are two lines of *B. rapa* ssp. *pekinensis* with distinct requirements for floral induction, requiring relatively short and long vernalization treatments, respectively. Rapid cycling *B. rapa* (RCBr), also known as Fast Plant, is a widely used model organism in biology education (Williams and Hill 1986; Musgrave 2000). RCBr was developed by selection of *B. rapa* for several traits, such as short time to flowering (16 days for flowering), rapid seed maturation, lack of seed dormancy, petite growth habit, and high female fertility (Williams and Hill 1986).

Several flowering-related markers have been identified from *B. rapa*, including a Random Amplified Polymorphic DNA (RAPD) marker associated with late-bolting in non-heading type Chinese cabbage (Ajisaka et al. 2001). Li et al. (2009) found some single nucleotide polymorphism (SNP) markers found in *B. rapa* *FLOWERING LOCUS C1* (*BrFLC1*) and *BrFLC2* correlated with differences in bolting time, while Yuan et al. (2009) developed cleaved amplified polymorphic sequences (CAPS) markers using SNPs found at the splice site of the *BrFLC1* gene. In addition, Wu et al. (2012) found that a naturally occurring insertion/deletion (InDel) variation in *BrFLC2* is associated with flowering time variation in *B. rapa*. Kakizaki et al. (2011) reported that QTLs associated with flowering time overlapped the positions of the *FLC* and *FLOWERING LOCUS T (FT)* genes, although molecular markers for these have not been developed yet. There are no reports to date concerning use of the core circadian clock gene, *CCA1*, to identify flowering time-related markers in *B. rapa*.

We provide evidence that sequence polymorphisms in the *B. rapa* *CCA1* gene (*BrCCA1*), particularly

those in the fourth intron, are linked to flowering time phenotypes, early vs. late flowering. We have confirmed its association using genomic PCR and high resolution melting (HRM). The high level of sequence polymorphisms in the fourth intron of *CCAI* is also found in many economically important *Brassica* crop plants and thus may be used to develop universal markers valuable for *Brassica* breeding programs.

Materials and methods

Plant materials

To analyze nucleotide polymorphisms in *CCAI* associated with variation in flowering, two inbred lines of *B. rapa* ssp. *pekinensis* with different flowering phenotypes, Chiifu and Kenshin (Li et al. 2013), and an RCB (IBM-218-DH1) were selected as reference lines. In addition, leaf samples for inbred lines of *B. rapa* were obtained from three different Korean seed companies: Hankook Seed Co. (indicated by CI or HS in line names), Nonghyup Seed Co. (indicated by BN or NS in line names) and Woori Seed Co. (indicated by WS in line names). *B. rapa* germplasm for flowering evaluation and marker analysis and additional *Brassica* species used in *CCAI* gene sequence analyses were selected from germplasm collection (Pang et al. 2015). For expression study of *BrCCAI*, reference plants were grown under long day (LD, 16h light and 8h dark) and short day (SD, 8h light and 16h dark) conditions until 21 day after germination, and sampled at indicated times.

Cloning and sequence analysis of *CCAI* genes

Genomic DNA was isolated from young leaves using the DNeasy Plant Mini kit (QIAGEN GmbH, Hilden, Germany). Information on the genomic DNA sequence of *Arabidopsis thaliana* *CCAI* (AT2G46830) was used as a reference for cloning *BrCCAI* genomic fragments from three reference lines and seven inbred lines (Fig. 1A). We compared the *A. thaliana* sequences with known *Brassica* *CCAI* sequences in the NCBI database (<http://ncbi.nlm.nih.gov>, as of August 2015) and designed four pairs of primers (Fig. 1A and Table 1). As shown in Fig. 1A, the universal primers (uCCA1F and uCCA1R) (Table 1) were designed

to amplify sequences between the start and stop codons from all *Brassica* species, using sequence information provided by the PCR products obtained with the four primer sets described above.

Genomic PCR was performed under the following conditions: denaturation (5 min at 94 °C), 30 cycles of amplification (30 s at 94 °C, 30 s at 52 °C, and 3 min at 72 °C), and a final extension (7 min at 72 °C). PCR products were purified using a MEGA-Spin gel extraction kit (Intron Biotech. Inc., Sungnam, Korea) and cloned into the TA-vector using the T&A cloning kit (RBC Bioscience Corp., New Taipei city, Taiwan). *Escherichia coli* (DH5α) cells were transformed with plasmid DNA carrying the desired insert. Plasmid DNA was purified using DNA-Spin (Intron Biotech. Inc., Sungnam, Korea) before sequencing. Since *Brassica* might contain multiple *CCA1* alleles, ten clones from each line were sequenced and analyzed. Any possible PCR and/or sequencing errors were eliminated by aligning the ten independent sequences obtained. All sequences were aligned on the published *Arabidopsis CCA1* sequence. Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used for multiple sequence alignment and neighbor-joining (NJ) method was used for the construction of phylogenetic trees.

RT-PCR analysis

Total RNA was isolated from samples using TRIzol reagent (Invitrogen, Carlsbad, USA), and further purified with a NucleoSpin RNA Clean-up Kit (Macherey-Nagel GmbH & Co., Düren, Germany). The purified RNA (1 µg) from each sample was synthesized using the ReverTraAce-α kit with the Oligo(dT) primers (Toyobo, Osaka, Japan). Complementary DNA was diluted 10-fold, and 1 µl of diluted cDNA was used in a 20 µl PCR mixture. RT-PCR primers for *BrCCA1* were 5'-CACTCCTTCGAGTAGTGACGATGTTG (forward) and 5'-TAGGATTGGTGTGCTGATGACTCTAC (reverse), and primers for *BrACT2*, used as a control, were 5'-GAACCGGGTGCTCCTCAGGA-3' (forward) and 5'-ATGGTACCGGAATGGTCAAGGC-3' (reverse). A standard RT-PCR was performed with a 5 min denaturation at 94 °C, followed by 25 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 90 s. PCR products were analyzed by electrophoresis using 1.2% agarose gels.

Development of InDel markers and optimization of PCR condition

Based on *CCA1* sequences covering the fourth exon, fourth intron, and a 5' part of the fifth exon (Fig. 1B; Supplementary Figs 1 and 2), three pairs of primers were designed to differentiate between the *B. rapa* reference lines RCBBr, Kenshin, and Chiifu (Table 1). In addition, control primers against the gene encoding 5.8S RNA were designed to confirm the presence of genomic DNA.

To optimize the PCR conditions for each primer pair (CCA1-F and CCA1-R1; CCA1-F and CCA1-R2; and CCA1-F and CCA1-R3), several annealing temperatures, extension times, and cycle numbers were tested and set. For the CCA1-F and CCA1-R1 primer pair, the PCR conditions were 5 min at 94 °C; followed by five cycles of 30 s at 94 °C, 30 s at 40 °C, and 30 s at 72 °C; then 25 cycles of 30 s at 94 °C, 30 s at 60 °C, and 30 s at 72 °C; with a final extension phase of 5 min at 72 °C. For CCA1-F and CCA1-R2 or CCA1-F and CCA1-R3 primer pairs, the reaction conditions were 5 min at 94°C, followed by 30 cycles of 30 s at 94 °C, 30 s at 55 °C, 30 s at 72 °C, and a final extension of 5 min at 72 °C. To confirm the presence of genomic DNA, 5.8S primers were mixed together with *CCA1* primers in each PCR reaction mix. PCR products produced by InDel markers were separated using 1.5% agarose gels for genotyping.

Flowering time evaluation and InDel marker analysis

To investigate the possible association between variations in *Brassica CCA1* sequences and flowering time, a total of 55 *B. rapa* accessions, including Chinese cabbage, pack choi, brown sarson type and caixin type, were used (Supplementary Table 1). The flowering time evaluation was carried out in a greenhouse and open field of Chungnam National University, Daejeon, Korean, from August to November, 2015. Germinated seeds were vernalized under 4°C for one month. Then six individuals for each accessions were sown in pots in glasshouse with 18h photoperiod, and in the open field, respectively. For the open field experiment, the vernalized seeds were sown in the cell tray, grown in a greenhouse for one month, and then transplanted to the open field. Flowering time was measured as days to flowering as described by Li et al. (2013).

To associate the flowering time variation with *BrCCA1* gene sequence variation, newly developed

InDel markers were applied to amplify the *CCAI* genomic DNA fragments in various *B. rapa* accessions. Two different types of PCR amplicons were detected by high resolution melting (HRM) method in a light Scanner System (Biofire Diagnostics, Salt Lake city, USA), as previously described by Li et al. (2013). The association between marker genotype and flowering time phenotype was tested by one way analysis of variance (ANOVA) using SPSS statistic package.

Results

Cloning and sequence analysis of *CCAI* genes from *B. rapa*

CCAI was selected as a candidate gene to develop flowering time markers for two reasons. Firstly, *CCAI* controls various plant growth and development including photoperiodic flowering as mentioned in introduction. Secondly, a previous study reported the overlap of many clock and flowering time QTLs (Lou *et al.* 2011). RCB_r is photoperiod-independent (Weis 2015) and requires only 16 days of growth for floral induction (Williams and Hill 1986). Although Kenshin and Chiifu require both long day conditions and vernalization for floral induction, only about 2 weeks of vernalization are necessary for Kenshin compared to more than 8 weeks of low temperature (~4 °C) treatment for Chiifu. Inbred lines provided by several Korean seed companies were found to be either EF (bolting and flowering 20-30 days earlier than LF) or LF (requiring growth in long day conditions after an extended period of vernalization) (Li et al., 2013 and Data not shown).

To investigate possible sequence variation in the *CCAI* gene of *B. rapa*, which could be used to develop flowering-time markers, we used genomic DNA isolated from three representative *B. rapa* lines (Chiifu, Kenshin, and RCB_r) and PCR primers designed based on the *Arabidopsis CCAI* sequence (Table 1 and Fig. 1) to clone *CCAI* genomic DNA fragments. The clones were sequenced and trimmed at the 5' and 3' untranslated regions (UTRs) in order to compare the amplified sequences between the start and stop codons with the *B. rapa CCAI* reference sequence.

Sequencing of ten individual *CCAI* clones from the three *B. rapa* inbred lines revealed that each of these inbred lines contained only one type of *CCAI* sequence. All *BrCCAI* genomic sequences were shorter

than the 2,713 base pair (bp) *Arabidopsis CCA1* sequence. The *CCA1* sequences amplified from RCBBr (2,510 bp) and Kenshin (2,516 bp) were about 150 bp longer than that obtained from Chiifu (2,367 bp). *CCA1* nucleotide sequences in RCBBr and Knshin showed 97.6% identity. Based on the size of *BrCCA1*, we considered RCBBr and Kenshin *CCA1* as EF type, while that of Chiifu as LF type (Williams and Hill 1986; Fig. 4A). The difference of *CCA1* sequence lengths between EF and LF lines was largely due to the length of the fourth intron, which was 380 bp long in the EF lines but 229 bp in the LF line (Fig. 1B and Supplementary Fig. 1).

To determine whether the polymorphisms in *BrCCA1*, particularly those around the fourth intron, could be used to predict flowering phenotypes, we compared the sequence similarities of *CCA1* from *Arabidopsis*, Chiifu, Kenshin, RCBBr, and additional *B. rapa* inbred lines (heading type Chinese cabbages), kindly provided by Korean seed companies, which displayed contrasting flowering phenotypes. As shown in Figure 2, the sequence similarities could be used to predict flowering behavior in *Brassica*. *CCA1* genomic DNA sequences obtained for current study were deposited in NCBI database (<http://www.ncbi.nlm.nih.gov>) with following accession numbers: KU201267 (*B. rapa* Chiifu), KU201266 (*B. rapa* Kenshin), KU201265 (*B. rapa* RCBBr), KU291166 (*B. rapa* BN515), KU291168 (*B. rapa* BN569), KU291164 (*B. rapa* BN615), KU291163 (*B. rapa* CI79), KU291167 (*B. rapa* CI8), KU291165 (*B. rapa* WSLF), KU291165 (*B. rapa* WSLF), and KX611131 (*B. rapa* WSEF). When a phylogenetic tree was constructed using genomic sequences from the second to fifth exons of *CCA1*, clear groupings reflecting the distinct flowering phenotypes were detected in the *B. rapa* lines (Fig. 2). The same pattern was observed when the analysis was repeated using the full length *CCA1* sequences between the start and stop codons (data not shown). Of the *BrCCA1* sequences, those from the fourth intron were found to be the most polymorphic, as they contained many InDels.

Development of flowering time markers

To develop InDel markers to quickly distinguish polymorphisms associated with the flowering time phenotypes of different *B. rapa* plants, we designed three primer sets against insertion/deletion sequences

in and around the fourth intron of *CCA1* (Table 1, Fig. 1B, Supplementary Figs. 1 and 2). CCA1-F, which binds to the fourth exon of *CCA1*, was a common forward primer and CCA1-R1 was a reverse primer targeted on the fifth exon. Depending on the source of the genomic DNA templates, PCR reactions using CCA1-F and CCA1-R1 produced DNA fragments of different sizes. A fragment of about 300 bp was amplified from the LF line, Chiifu, whereas fragments of approximately 450 bp were produced from the EF lines, Kenshin and RCBBr (Fig. 3A).

Two other reverse primers, CCA1-R2 and CCA1-R3, were designed to bind to regions of the fourth intron containing the InDel and to be used with CCA1-F to produce PCR products only in the EF reference *B. rapa* plants (Table 1, Fig. 1B, Supplementary Figs. 1 and 2). The performance of these primers was tested using various *B. rapa* inbred lines (Fig. 3B and 3C). When a combination of the CCA1-F and CCA1-R2 primers was used, a clear amplification product for *CCA1* was detected in EF inbred lines, but weak amplification was also observed in several LF lines. We concluded, therefore, that this primer set was not useful for predicting flowering time phenotype in *B. rapa*. By contrast, the combination of CCA1-F and CCA1-R3 primers resulted in a clear difference between PCR products that reflected flowering type. *CCA1* PCR products were amplified only from the EF inbred lines, indicating that this primer set was a reliable predictor of flowering time behavior in *B. rapa*. It is worth noting that PCR with the CCA1-F and CCA1-R1 primers was necessary to determine whether plants contained the Chiifu-type *CCA1* gene associated with LF.

Whether InDel in the fourth intron of *CCA1* affects expression of the gene, three reference lines grown in LD and SD condition were subjected to RT-PCR experiment (Fig. 3D). The expression pattern under SD condition was same in all three lines, while transcript was detected in the middle of long day period in both RCBBr and Kenshin. Although *CCA1* expression pattern could not be directly related to flowering phenotypes, InDel might be associated with flowering phenotype with unknown mechanism.

Association between flowering time variation to BrCCA1 sequence variation among *B. rapa*

We investigated flowering time among 55 *B. rapa* accessions including 48 Chinese cabbage, 3 Pak

choi, 3 caixin and 1 sarson cultivars selected from germplasm collection (Pang et al. 2015) in two independent experiment. Experiment I was performed in a greenhouse with artificial control of 18 h photoperiod, and experiment II in the open field. The flowering time varied from 23 to 90 days with average 39.3 days from the seeds germination to the opening of the first flower in greenhouse experiment, and 23 to 95 with average 42.2 days in the open field experiment. Two caixin cultivars flowered 23 days and one cultivars flowered 32 days after germination in both experiments. Brown sarson type flowered 23 days in both experiments. Pak choi types flowered from 23 to 32 days in greenhouse, while flowered from 23 to 40 days in open field.

We screened *BrCCA1* InDel marker over the 55 *B. rapa* fixed lines by using HRM technology. Among three primer sets, only CCA1-F/R1 performed well and can be detected by HRM method. The grey and red curves represented the deletion allele and insertion allele, respectively (Fig. 4). A total of 31 accessions were found to have insertion allele, including 24 Chinese cabbage type, all three pack choi, three caixin, and one sarson types. The deletion allele was present in 24 Chinese cabbage accessions. The correlation analysis showed that the *BrCCA1* InDel marker polymorphism was significantly associated with the variation of flowering time among 55 *B. rapa* accessions in both greenhouse and open field experiment (Table 2). Accessions with the *BrCCA1* insertion allele flowered early with an average of 31.8 and 33.4 days in greenhouse and open field experiment, respectively. While, the accessions with the *BrCCA1* deletion allele flowered late with an average of 48.8 days in greenhouse experiment, and 53.4 days in the open field (Table 2).

Polymorphisms of *CCA1* in *Brassica* species

Sequence variation in *CCA1*, similar to that identified during the marker development for *B. rapa*, is also present in other economically important *Brassica* species, such as *B. napus* and *B. oleracea*. *CCA1* genes appear to have been preferentially retained in *B. rapa* during the diploidization process (Lou et al. 2012), which followed separation from the common ancestor with *Arabidopsis* and whole genome triplication. Therefore, we also cloned and analyzed *CCA1* gene sequences from other *Brassica* plants

found in U's Triangle (U 1935). We cloned at least 10 clones from each plant and representative sequences were deposited in the NCBI with sequence accession numbers as follows: KU292582 (*B. rapa* Caixin), KU292578 (*B. rapa* PakChoi-1), KU292579 (*B. rapa* PakChoi-2), KU292580 (*B. nigra* BB-1), KU292581 (*B. nigra* BB-2), KU292583 (*B. oleracea* CC-1), KU292584 (*B. oleracea* CC-2), KU292596 (*B. oleracea* RCBo) KU292585 (*B. oleracea* CC-F1-1), KU292586 (*B. oleracea* CC-F1-2), KU292587 (*B. juncea* AABB-1), KU292588 (*B. juncea* AABB-2), KU292589 (*B. juncea* AABB-3), KU292590 (*B. napus* AACC-1), KU292591 (*B. napus* AACC-2), KU292592 (*B. napus* AACC-3), KU292593 (*B. carinata* BBCC-1), KU292594 (*B. carinata* BBCC-2), KU292595 (*B. carinata* BBCC-3).

We found that short fourth introns (228 or 229 bp) were only found in plants containing the A genome, while clearly longer fourth introns (310 to 479 bp) were detected in the B and C genomes, in addition to EF type of *B. rapa* plants carrying the A genome (Fig. 5). Each inbred line contained only a single type of *CCAI*, but many cultivars contained more than one *CCAI* variant. All but one of the allotetraploids contained two different types of *CCAI* originating from two different genomes, as expected for genome hybridization (Fig. 5). The C genome-type *CCAI*, however, was not identified in the BBCC genome of *B. carinata*, indicating an asymmetric origin of *CCAI* from the B genome in this species. Instead, *B. carinata* contained shorter versions of the B genome-type *CCAI* fourth intron (310 bp) together with a regular B genome-type fourth intron sequence (406 bp). The sequences of the fourth intron of *CCAI* in the BBCC-2 and BBCC-3 genomes of *B. carinata* were found to be identical and generated by a 96 bp deletion from BBCC-1 (Fig. 5). The size of *CCAI* was largely determined by the size of the fourth intron, although the seventh exon made a minor contribution to size. When all the *CCAI* in Figure 5 were clustered according to their genomic sequence similarities, three clades were noticed (Fig. 6A). All *B. rapa* *CCAI* sequences were found in two different clades, which were closely associated with each other but separated from a third clade containing *Arabidopsis* *CCAI*. Two clades, consisting of *Brassica* *CCAI*, appeared to be determined mainly by fourth intron sequences, since a distinct but similar grouping pattern was observed when only sequences from the fourth intron of *CCAI* were considered (Fig. 6B). AABB-2, AABB-3, and AACC-3, which all contain a Chiifu-type fourth intron, were grouped together. If *CCAI* variants with a Chiifu-type

fourth intron also determine the LF phenotypes of allotetraploid species, the markers developed here could also be utilized in breeding programs for many other *Brassica* species.

Discussion

Flowering time markers that distinguished between EF and LF phenotypes of *B. rapa* were developed based on sequence polymorphisms in the *B. rapa CCA1* gene. CCA1 is a well-known core component of the circadian clock, which controls plant metabolism and life cycle. CCA1 regulates photoperiodic flowering through transcriptional regulation of flowering pathway genes (Niwa *et al.* 2007; Fujiwara *et al.* 2008). *CCA1* expression is responsive to environmental stresses and regulated by alternative splicing (Seo *et al.* 2012; Filichkin *et al.* 2015). The involvement of *CCA1* in the flowering time of plants has been suggested by several research groups including Slotte *et al.* (2007) (Imaizumi 2010; Li *et al.* 2011; Lou *et al.* 2011; Lu *et al.* 2012). Recent reports that CCA1 binds to thousands of genes (Nagel *et al.* 2015; Kamioka *et al.* 2016) imply CCA1 plays an important role in regulation of a myriad of biological processes and stress responses.

Flowering time is one of the most important traits of commercial varieties of *B. rapa* and, therefore, molecular markers predicting flowering time phenotypes are invaluable in breeding programs. A few flowering time-related markers have been developed previously for *B. rapa*, including a CAPS marker using SNPs found at the splice site of *BrFLC1* (Yuan *et al.* 2009) and InDel markers from *BrFLC2* (Wu *et al.* 2012). However, no flowering time marker using polymorphisms in *B. rapa CCA1* has so far been developed. Although DNA sequence polymorphisms in *CCA1* genes were found in all the introns and exons investigated, the most dramatic differences were detected in the fourth intron. Most importantly, this variation was perfectly associated with flowering time behavior (EF or LF) of the various *B. rapa* lines used in this study (Figs 2-4). The best PCR-based marker was CCA1-F/CCA1-R3 combination (Fig. 3), while CCA1-F/CCA1-R1 combination for HRM confirmation (Fig. 4). Using CCA1-F/CCA1-R1 for PCR, we could discriminate homozygote and heterozygote with respect to InDel expressing flowering phenotypes (data not shown). Insertion in fourth intron seemed to be dominant against deletion allele associated with LF phenotype. Taken together, the PCR-based markers and HRM amplicon we have

13

developed based on this variation will be useful to the *B. rapa* breeding program.

Regarding to *Brassica* species in U's triangle, the deletion in the fourth intron seemed to be originated from LF type AA genome. We are not sure whether deleted 4th intron reflects LF phenotype in AABB and AACC. Interestingly, BBCC genome contained intermediate type 4th intron. Recently, Schiessl *et al.* (2015) reported that an SNP found in *B. napus CCA1* (*BnCCA1*) is associated with flowering time. This indicates that the considerable variation found in the fourth intron sequences of *CCA1* across *Brassica* species provides valuable information on the evolution and breeding history of these crop plants. In addition to the 4th intron polymorphism, large number of InDel and SNPs were found in exons as well as introns of *CCA1* in *Brassica* CC genomes, which contain similar size of genome. These also imply that sequence polymorphisms found in *Brassica CCA1* gene could be associated with various traits, thereby they can be used in various molecular markers.

The possible functions of the sequence variation observed in the fourth intron of *Brassica CCA1* remain to be determined. Palmer and Logsdon (1991) have speculated that the insertion of introns into genes occurred late in eukaryote evolution in order to regulate expression. Intron-mediated enhancement of gene expression (Gallegos and Rose 2015) and alternative splicing of transcripts in response to abiotic conditions (Seo *et al.* 2012; Cui *et al.* 2014; Lunghi *et al.* 2015) have been documented, and intron retention is frequently associated with abiotic stress conditions in barley (Panahi *et al.* 2015) and tumor-suppressor inactivation in animals (Jung *et al.* 2015). Gene regulation by intronic miRNA, generated by the insertion of transposons, is reported from a variety of animals (Ying *et al.* 2010 and 2013). In a similar manner, the dramatic variations found in and around the fourth intron of *Brassica CCA1* may affect agronomically important traits, such as flowering time, by regulating the expression of *CCA1*. However, our expression study of *BrCCA1* (Fig. 3D) did not support the regulation of mRNA expression by intron variation under the normal growth condition. However, we could not rule out possibility of differential expression of *BrCCA1* under particular conditions. Given that a functional role in cold acclimation for alternatively spliced transcripts of *CCA1*, which retain the fourth intron, has been established in *Arabidopsis* (Seo *et al.* 2012), it should be determined whether alternative splicing of *CCA1* also occurs in

Brassica and, if so, whether this affects the performance of crops.

Our results provide an important first step in the development of universal *Brassica* InDel markers that can predict a desired phenotype, in keeping with recent work in many crop species (Salgotra *et al.* 2014; Kage *et al.* 2015;).

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by a grant from the Technology Development Program for Agriculture and Forestry, Ministry for Food, Agriculture, Forestry, and Fisheries (Grant 607003-05-5-SB310), Republic of Korea.

References

- Ajisaka H., Kuginuki Y., Yui S., Enomoto S., and Hirai M. 2001. Identification and mapping of a quantitative trait locus controlling extreme late bolting in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis* syn. *campestris* L.) using bulked segregant analysis. *Euphytica* **118**: 75-81. doi:10.1023/A:1004023532005
- Cockram J., Jones H., Leigh F.J., O'Sullivan D., Powell W., Laurie D.A., and Greenland A.J. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *J. Exp. Bot.* **58**: 1231-1244. doi: 10.1093/jxb/erm042
- Cui Z., Xu Q., and Wang X. 2014. Regulation of the circadian clock through pre-mRNA splicing in *Arabidopsis*. *J. Exp. Bot.* **65**: 1973-1980. doi: 10.1093/jxb/eru085
- Dong M.A., Farré E.M., and Thomashow M.F. 2011. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA*. **108**: 7241-7246. doi: 10.1073/pnas.1103741108
- Filichkin S.A., Cumbie J.S., Dharmawardhana P., Jaiswal P., Chang J.H., Palusa S.G., Reddy A.S., Megraw M., and Mockler T.C. 2015. Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in *Arabidopsis*. *Mol. Plant* **8**: 207-227. doi: 10.1016/j.molp.2014.10.011.
- Fujiwara S., Oda A., Yoshida R., Niinuma K., Miyata K., Tomozoe Y., Jajima T., Nakagawa M., Hayashi K., Coupland G., and Mizoruchi T. 2008. Circadian clock proteins LHY and CCA1 regulate SVP protein accumulation to control flowering in *Arabidopsis*. *Plant Cell* **20**: 2960-2971. doi: 10.1105/tpc.108.061531.
- Gallegos J.E., and Rose A.B. 2015. The enduring mystery of intron-mediated enhancement. *Plant Sci* **237**: 8-15. doi: 10.1016/j.plantsci.2015.04.017.
- Green R.M., and Tobin E.M. 1999. Loss of the circadian clock-associated protein 1 in *Arabidopsis* results in altered clock-regulated gene expression. *Proc. Natl. Acad. Sci. USA* **96**: 4176-4179. doi: 10.1073/pnas.96.7.4176
- Green R.M., and Tobin E.M. 2002. The role of CCA1 and LHY in the plant circadian clock. *Dev. Cell* **2**: 516-518. doi: 10.1016/S1534-5807(02)00184-3

- Imaizumi T. 2010. *Arabidopsis* circadian clock and photoperiodism: time to think about location. *Curr. Opin. Plant Biol.* **13**: 83–89. doi: 10.1016/j.pbi.2009.09.007.
- Jung C., and Müller A.E. 2009. Flowering time control and applications in plant breeding. *Trends. Plant Sci.* **14**: 563-573. doi: 10.1016/j.tplants.2009.07.005.
- Jung H., Lee D., Lee J., Park D., Kim Y.J., Park W.Y., Hong D., Park P.J., and Lee E. 2015. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47**: 1242-1248. doi: 10.1038/ng.3414.
- Kage U., Kumar A., Dhokane D., Karre S., and Kushalappa A.C. 2016. Functional molecular markers for crop improvement. *Crit. Rev. Biotechnol.* **36**: 917-930. doi: 10.3109/07388551.2015.
- Kamioka M., Takao S., Suzuki T., Taki K., Higashiyama T., Kinoshita T., and Nakamichi N. 2016. Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the *Arabidopsis* circadian clock. *Plant Cell* **28**: 696-711. doi: 10.1105/tpc.15.00737.
- Kakizaki T., Kato T., Fukino N., Ishida M., Hatakeyama K., and Matsumoto S. 2011. Identification of quantitative trait loci controlling late bolting in Chinese cabbage (*Brassica rapa* L.) parental line Nou 6 gou. *Breed. Sci.* **61**: 151-159.
- Lai A.G., Doherty C.J., Mueller-Roeber B., Kay S.A., Schippers J.H., and Dijkwel P.P. 2012. CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proc. Natl. Acad. Sci. USA* **109**: 17129-17134. doi: 10.1073/pnas.1209148109.
- Li F., Kitashiba H., Inaba K., and Nishio T. 2009. A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. *DNA Res.* **16**: 311-323. doi: 10.1093/dnares/dsp020
- Li G., Siddiqui H., Teng Y., Lin R., Wan X.Y., Li J., Lau O.S., Ouyang X., Dai M., Wan J., Devlin P.F., Den X.W., and Wang H. 2011. Coordinated transcriptional regulation underlying the circadian clock in *Arabidopsis*. *Nat. Cell Biol.* **13**: 616-622. doi:10.1038/ncb2219
- Li X., Ramchiary N., Dhandapani V., Choi S.R., Hur Y., Nou I.S., Yoon M.K., and Lim Y.P. 2013. Quantitative trait loci mapping in *Brassica rapa* revealed the structural and functional conservation of

genetic loci governing morphological and yield component traits in the A, B, and C subgenomes of *Brassica* species. *DNA Res.* **20**: 1-16. doi: 10.1093/dnares/dss029.

Lou P., Wu J., Cheng F., Cressman L.G., Wang X., and McClung C.R. 2012. Preferential retention of circadian clock genes during diploidization following whole genome triplication in *B. rapa*. *Plant Cell* **24**: 2415-2426. doi: 10.1105/tpc.112.099499.

Lou P., Xie Q., Xu X., Edwards C.E., Brock M.T., Weinig C, and McClung C.R. 2011. Genetic architecture of the circadian clock and flowering time in *Brassica rapa*. *Theor. Appl. Genet.* **123**: 397-409. doi: 10.1007/s00122-011-1592-x.

Lu S.X., Webb C.J., Knowles S.M., Kim S.H., Wang Z.Y., and Tobin E.M. 2012. CCA1 and ELF3 interact in the control of hypocotyl length and flowering time in *Arabidopsis*. *Plant Physiol.* **158**: 1079-1088. doi: 10.1104/pp.111.189670.

Lunghi M., Spano F., Magini A., Emiliani C., Carruthers V.B., and Di Cristina M. 2016. Alternative splicing mechanisms orchestrating post-transcriptional gene expression: intron retention and the intron-rich genome of apicomplexan parasites. *Curr. Genet.* **62**: 31-38. doi: 10.1007/s00294-015-0506-x.

McClung C.R. 2014. Wheels within wheels: New transcriptional feedback loops in the *Arabidopsis* circadian clock. *F1000Prime Rep.* **6**:2. doi: 10.12703/P6-2

Nagel D.H., Doherty C.J., Pruneda-Paz J.L., Schmitz R.J., Ecker J.R., and Kay S.A. 2015. Genome-wide identification of CCA1 targets uncovers an expanded clock network in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **112**: E4802-4810. doi: 10.1073/pnas.1513609112.

Niwa Y., Ito S., Nakamichi N., Mizoguchi T., Niinuma K., Yamashino T., and Mizuno T. 2007. Genetic linkages of the circadian clock-associated genes, *TOC1*, *CCA1* and *LHY*, in the photoperiodic control of flowering time in *Arabidopsis thaliana*. *Plant Cell Physiol.* **48**: 925-937. doi:10.1093/pcp/pcm067.

Palmer J.D., and Logsdon J.M. Jr. 1991. The recent origins of introns. *Curr. Opin. Genet. Dev.* **1**: 470-477.

Panahi B., Mohammadi S.A., Khaksefidi R.E., Fallah J., and Ebrahimie E. 2015. Genome-wide analysis of alternative splicing events in *Hordeum vulgare*: Highlighting retention of intron-based splicing and its possible function through network analysis. *FEBS Lett.* **589**: 3564-3575. doi: 10.1016/j.febslet.2015.09.023.

- Pang W., Li X., Choi S.R., Dhandapani V., Im S., Park M.Y., Jang C.S., Yang M.S., Ham I.K., Lee E.M., Kim W., Lee S., Bonnema G., Park S., Piao Z., and Lim Y.P. 2015. Development of a leafy *Brassica rapa* fixed line collection for genetic diversity and population structure analysis. *Mol. Breeding* **35**:54. DOI: 10.1007/s11032-015-0221-9
- Salgotra R.K., Gupta B.B., and Stewart C.N. Jr. 2014. From genomics to functional markers in the era of next-generation sequencing. *Biotechnol. Lett.* **36**: 417-426. doi: 10.1007/s10529-013-1377-1.
- Schiessl S., Iniguez-Luy F., Qian W., and Snowdon R.J. 2015. Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*. *BMC Genomics* **16**: 737. doi: 10.1186/s12864-015-1950-1.
- Seo P.J., Park M.J., Lim M.H., Kim S.G., Lee M., Baldwin I.T., and Park C.M. 2012. A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in *Arabidopsis*. *Plant Cell* **24**: 2427-2442. doi: 10.1105/tpc.112.098723.
- Slotte T., Holm K., McIntyre L.M., Lagercrantz U., and Lascoux M. 2007. Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (Brassicaceae). *Plant Physiol.* **145**: 160-173. doi: [10.1104/pp.107.102632](https://doi.org/10.1104/pp.107.102632).
- Song Y.H., Ito S., and Imaizumi T. 2013. Flowering time regulation: photoperiod- and temperature sensing in leaves. *Trends Plant Sci.* **18**: 575-583. doi: 10.1016/j.tplants.2013.05.003.
- Song Y.H., Shim J.S., Kinmonth-Schultz H.A., and Imaizumi T. 2015. Photoperiodic flowering: time measurement mechanisms in leaves. *Annu. Rev. Plant Biol.* **66**: 441-464. doi: 10.1146/annurev-arplant-043014-115555.
- Srikanth A., and Schmid M. 2011. Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* **68**: 2013-2037. doi: 10.1007/s00018-011-0673-y.
- U N. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japan. J. Bot.* **7**:389-452.
- Wang N., Qian W., Suppanz I., Wei L., Mao B., Long Y., Meng J., Müller A.E., and Jung C. 2011. Flowering time variation in oilseed rape (*Brassica napus* L.) is associated with allelic variation in the

- FRIGIDA* homologue *BnaA.FRI.a*. J. Exp. Bot. **62**: 5641-5658. doi: 10.1093/jxb/err249.
- Wang Z.Y., Kenigsbuch D., Sun L., Harel E., Ong M.S., and Tobin E.M. 1997. A Myb-related transcription factors is involved in the phytochrome regulation of an *Arabidopsis Lhcb* gene. Plant Cell **9**:491-507. doi: 10.1105/tpc.9.4.491.
- Wang Z.Y., and Tobin E.M. 1998. Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* gene disrupts circadian rhythms and suppresses its own expression. Cell **93**: 1207-1217. doi: 10.1016/S0092-8674(00)81464-6
- Weis A.E. 2015. Inheritance of rapid cycling in *Brassica rapa* fast plants: dominance that increases with photoperiod. Int. J. Plant Sci. **176**: 859-868. doi: 10.1086/683304.
- Williams P.H., and Hill C.B. 1986. Rapid-cycling populations of *Brassica*. Science **232**: 2385-1389. doi: 10.1126/science.232.4756.1385.
- Wu J., Wei K., Cheng F., Li S., Wang Q., Zhao J., Bonnema G., and 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. doi: 10.1186/1471-2229-12-151.
- Ying S.Y., Chang C.P., and Lin S.L. 2010. Intron-mediated RNA interference, intronic microRNAs, and applications. Methods Mol. Biol. **629**: 205-237. doi: 10.1007/978-1-60761-657-3_14.
- Ying S.Y., Chang D.C., and Lin S.L. 2013. The MicroRNA. Methods Mol. Biol. **936**: 1-19. doi: [10.1007/978-1-62703-083-0_1](https://doi.org/10.1007/978-1-62703-083-0_1)
- Yuan Y.X., Wu J., Sun R.F., Zhang X.W., Xu D.H., Bonnema G., and Wang X.W. 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. doi: 10.1093/jxb/erp010.
- Zhao J., Wang X., Deng B., Lou P., Wu J., Sun R., Xu Z., Vromans J., Koornneef M., and Bonnema G. 2005. Genetic relationships within *Brassica rapa* as inferred from AFLP fingerprints. Theor. Appl. Genet. **110**: 1301-1314. doi: 10.1007/s00122-005-1967-y.

Table 1 Names and sequences of primers used in *CCA1* gene cloning and marker development.

Forward Primer		Reverse Primer		Remarks
Name	Sequence	Name	Sequence	
FP1	ATGGAGACWAATTCGTCTGGAGAA	RP1	TACCTTGGAGAAAAATTCTGAGCRTG	Exon 1-4
FP2	GTAGCAACGAAAAGTCTGTCCAGAT	RP2	CCAAGGGACTCTTTCCATCATTC	Exon 4-5
FP3	GTAGAGAAAGAGGCTGAAGCTAAAGGTGT A	RP3	YGCCTCAACATCGTCACTACTCGA	Exon 5-7
FP4	TTGCAGCGCWAGTGCTTG	RP4	TCAAAYGTTACARGAAGACTATGGACAA	Exon 5-8
uCCA1F	ATGGAGACWAATTCGTCTGGAGAAG	uCCA1R	TCATGTKGAAGYTTGWGTTTCCAAC	ATG to TGA 298 bp for LF
CCA1-F	GTCCAGATTAGAAGCCATGCTCAGA	CCA1-R1	GTACACCTTTAGCTTCAGCCTCTTTCTC	450 bp for EF
		CCA1-R2	TGTTACCTCTGTGGCTCATCATCAC	329 bp for EF
		CCA1-R3	GTTACCCTCTAGATAAAACACAAACAAACAT ATAT	395 bp for EF
5.8S-F	ACGACTCTCGGCAACGG	5.8S-R	GGGCGCAACTGCGTTCAAA	

EF and LF indicate early and late flowering lines, respectively.

Table 2 Average flowering time in different *B. rapa* accessions according to the *BrCCA1* genotype.

Genotype	No. of lines	FT in greenhouse	FT in open field
Deletion	24	48.8±2.5a*	53.4±1.9a
Insertion	31	31.8±0.7b	33.4±1.6b

*Significance was determined by one way ANOVA at the $p<0.01$ level.

FT: number of days to flowering

Draft

Figure Captions

Fig. 1 Structure of the *CCA1* gene indicating the positions of primers used in cloning of *CCA1* from *Brassica* and markers to predict flowering time. **A**, *Arabidopsis thaliana* *CCA1* (*AtCCA1*) gene structure and corresponding positions of *Brassica* primers used in this study. Primers were designed by comparing *AtCCA1* with *CCA1* sequences known from *Brassica* species. Boxes depict exons: open boxes indicate untranslated regions; filled boxes indicate coding sequences. FP: forward primer; RP: reverse primer; E: exon. uCCA1f and uCCA1r indicate universal primers used in the cloning of *CCA1* from all *Brassica* species. PCR using universal primers amplified full length *CCA1* genomic DNA between the start (ATG) and stop (TGA) codons. **B**, Representative structures of *CCA1* genes from *B. rapa* inbred lines showing the positions of primers used in flowering time marker development. *CCA1* from RCBBr (2,510 bp) and Kenshin (2,516 bp) showed 97.6% sequence identity. *CCA1* from Chiifu (2,367 bp) was 149 bp shorter than that from Kenshin. Areas within dotted lines include the fourth exons, fourth introns, and fifth exons.

Fig. 2 Clustering of *B. rapa* inbred lines based on *CCA1* sequence similarities. Genomic DNA sequences from exons 2 to 5 of *CCA1* (see Supplementary Fig. 2), including intron sequences, were used in the analysis with ClustalW2. *CCA1* from *A. thaliana* was included as an outgroup. Inbred lines, whose names start with CI, BN, and WS, were established by Hankook Seed Co., Nonghyup Seed Co., and Woori Seed Co, respectively. Early (EF) and late flowering (LF) *B. rapa* inbred lines are indicated on the right hand side.

Fig. 3 PCR products generated using three primer sets and expression of *BrCCA1*. **A**, PCR products of reference lines using CCA1F/CCA1-R1 primers. Arrows indicate DNA fragments amplified from *CCA1* differing between early and late flowering inbred lines of *B. rapa*. **B**, PCR products amplified from various *B. rapa* inbred lines by CCA1-F and CCA1-R2 primers. Multiplex PCR was performed using primer pairs for *CCA1* and 5.8S. The latter was included as internal control to monitor the quantity and quality of the

genomic DNA template and proper PCR amplification. F_1 indicates the progeny from a Chiifu $\times$ Kenshin cross. The letters 'WS', 'NS', and 'HS' in the names of inbred lines refer to the three Korean seed companies that provided the leaf samples; the letters 'EF' and 'LF' indicate early and late flowering behavior, respectively. Late flowering *B. rapa* lines are underlined. **C**, PCR products amplified from various *B. rapa* inbred lines using CCA1-F and CCA1-R3 primers. PCR-condition and abbreviations are same as B. **D**, Expression of *BrCCA1* gene under long day and short day conditions. Plants were grown in LD or SD for 3 weeks and sampled at indicated times.

Fig. 4 Confirmation of flowering phenotype by HRM analysis. **A**, Amplicon scanning analysis of reference lines 'Chiifu', 'Kenshin', and 'RCBr' amplified by CCA1F/R1 primer set using light scanner system. The different curves show deletion alleles of 'Chiifu' type in red and insertion alleles of 'Kenshin' and 'RCBr' in grey color. **B**, Amplicon scanning analysis of 55 *B. rapa* accessions screened by CCA1F/R1 primer. Red and grey curves represent *BrCCA1* deletion and insertion alleles found in early and lately flowering accessions, respectively.

Fig. 5 Exon/intron structures of *CCA1* genes from various *Brassica* species. Filled boxes and connecting lines indicate exons and introns, respectively. The gene sequence from the start codon in exon 2 to the stop codon in the eighth or ninth exon is displayed. The number above the fourth intron indicates its length in base pairs. The genome composition of each *Brassica* species or variety, including allotetraploids, is shown on the left-hand.

Fig. 6 Cluster analysis of *CCA1* gene sequences from various diploid and allotetraploid *Brassica* plants. The phylogenetic trees were constructed using (A) full genomic sequences between the start and stop codons or (B) fourth intron sequences only.

Fig1

124x86mm (300 x 300 DPI)

Fig2

59x46mm (300 x 300 DPI)

Color_Fig4

93x102mm (300 x 300 DPI)

<https://mc06.manuscriptcentral.com/genome-pubs>

64x49mm (300 x 300 DPI)

Supplementary Table 1 *B. rapa* collection used to confirm the association with *BraCCA1* sequence variation

Accession name	Cultivar group	Scientific name	Type
Chiifu	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
Kenshin	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
HKC-005	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
11Su-4	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
11Su-5	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
11Su-8	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
11Su-12	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
B2	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C1	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C2	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
E1	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
E2	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
ZS072518	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
501	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
FT-50	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
Tro-St-A	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
Tro-Jb	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C-25	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
09-FK15	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
09-FK18	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
Haek Baechu	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
Songdo Shin 2 Ho	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
CR-GJ-1	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
CR-GJ-2	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
GWGP	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
CHBC	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
CG2	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line

OHCR	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C-09	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C-16	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C-21	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C-22	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
C-34	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
HKC-001	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
HKC-002	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
HKC-003	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
HKC-006	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
S538-C	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
S55	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
S50-A	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	Inbred line
26110-87	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
26128-4	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
27160-16	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
27295-12	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
26130-2	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
26148-1	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
27143-26	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
26110-95	Chinese cabbage	<i>B. rapa</i> L. ssp. <i>pekinensis</i>	DH
CNU_28064	Caixin	<i>B. rapa</i> L. ssp. <i>parachinensis</i>	Inbred
27148	Caixin	<i>B. rapa</i> L. ssp. <i>parachinensis</i>	Inbred
Z062044	Caixin	<i>B. rapa</i> L. ssp. <i>parachinensis</i>	Inbred
28732	Brown sarson	<i>B. rapa</i> L. ssp. <i>sarson</i>	CV
CNU_10007	Pak choi	<i>B. rapa</i> L. ssp. <i>chinensis</i>	CV
25094	Pak choi	<i>B. rapa</i> L. ssp. <i>chinensis</i>	CV
Aijiaohuang	Pak choi	<i>B. rapa</i> L. ssp. <i>chinensis</i>	CV

Supplementary Fig. 1 *CCA1* sequences from the RCB_r, Kenshin, and Chiifu inbred lines of *B. rapa*. The entire genomic DNA sequence from the start (ATG) to the stop (TGA) codon is shown. Shaded regions indicate exons. The positions of primers used in marker development and their binding sites in reference lines are indicated. Forward primer: CCA1-F; reverse primers: CCA1-R1, CCA1-R2, and CCA1-R3.

Supplementary Fig. 2 Alignment of *CCA1* sequences (second to fifth exons, together with their intervening introns) from reference and inbred lines of *B. rapa*. Exon sequences are shaded in gray. The primers used in marker development, together with their binding sites on each *B. rapa* sequence, are indicated. The ten *Brassica* lines shown here are the same as those shown in Figure 2.

Draft

234x424mm (300 x 300 DPI)

96x53mm (300 x 300 DPI)