Conservation and function of *COOLAIR* long non-coding RNAs in *Brassica* flowering time control

Emily Jane Hawkes

Thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

John Innes Centre

September 2017

This page is intentionally left blank.

Abstract

Since their discovery long non-coding RNAs (lncRNAs) have in turn been described as essential genomic regulators or as transcriptional noise. Examples of lncRNAs with experimentally-validated function are limited, with poor nucleotide sequence conservation calling apparent functionality into question. *COOLAIR* lncRNAs are transcribed in the antisense direction at the *Arabidopsis thaliana* floral repressor gene, *Flowering Locus C* (*FLC*). Previous work has revealed a role for *COOLAIR* antisense RNAs in regulation of the *FLC* protein-coding sense transcript and, consequently, flowering time.

FLC homologues are wide-spread in flowering plants, but nucleotide sequence conservation of *COOLAIR*-specific regions is low. *COOLAIR* has a complex secondary structure, and covariant base-pair mutations predict strong conservation of this secondary structure across flowering plants. Syntenic transcription of *COOLAIR* was confirmed *in vivo* for several species within the family Brassicaceae, including three commercially important *Brassica* crops: *B. rapa*, *B. oleracea* and *B. napus*. *COOLAIR* transcription was detected from at least three of four ancient *FLC* clades within the latter three polyploid and paleopolyploid species. Each *FLC* homologue has distinct nucleotide sequence, expression patterns, and *COOLAIR* isoforms. Further variation in *COOLAIR* was identified between winter and spring cultivars. It was hypothesised that this could affect *cis*-regulation of *FLC*. Correlation between *COOLAIR* modifications and flowering time was tested by introducing antisense splice site mutations into *Brassica FLC* transgenes in *A. thaliana*. These experiments suggested that a shorter *COOLAIR* isoform with a disrupted structural motif was a weaker negative regulator of *FLC*.

This work supports conservation of *COOLAIR* expression, structure and function in *Brassica* crop plants, and a role for RNA structure in *COOLAIR* function. We propose an evolutionarily conserved lncRNA that is neither essential regulator nor transcriptional noise, but rather adapts with the environment to fine-tune the transition to flowering.

Acknowledgements

First and foremost, I would like to thank my supervisors, Judith Irwin and Caroline Dean, for their support and guidance over the last four years. Thank you not only for helping to develop and progress this project, but for your encouragement and enthusiasm throughout.

I acknowledge the BBSRC and my iCASE partner, Bayer CropScience, for funding this work. Thank you to my iCASE supervisors, Chikako Shindo, Pieter Ouwerkerk and Mark Davey, for your guidance and scientific support. Further, thank you to the above and to the Genetics team for making me feel so welcome during my visits to the Gent site in Belgium, and to Freya Lammertyn for technical assistance during my first placement.

I acknowledge our collaborators at Los Alamos National Laboratory, New Mexico, U.S.A. Thank you to Karissa Sanbonmatsu, Scott Hennelly, Irina Novikova and team for introducing me to the chemical-probing of RNA secondary structures, for solving the *in vitro COOLAIR* secondary structures, and for working with us to produce a paper in this field. Thank you, Karissa, for many fascinating discussions on secondary structure and for explaining your methods of secondary structure prediction and covariation analysis.

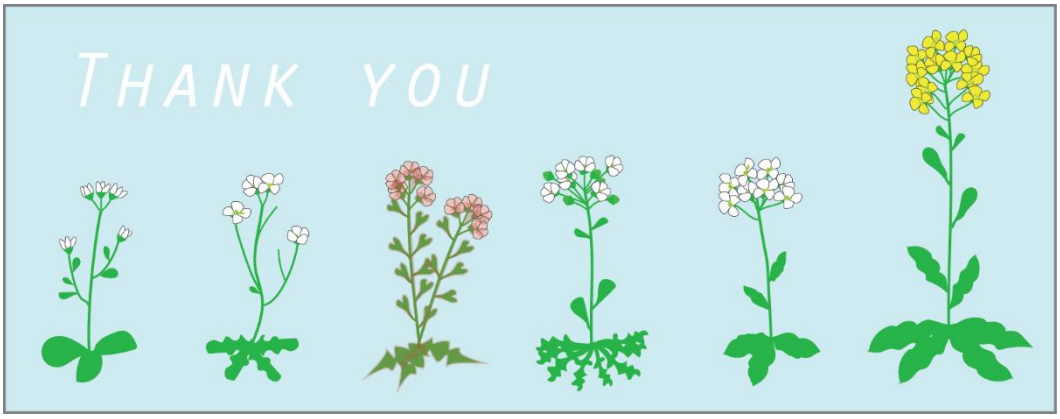
I thank members of the Dean lab, past and present, who have provided technical assistance, project advice and many useful discussions. Thank you especially to Qianwen Sun for your supervision at the beginning of this project and to Zhe Wu for technical assistance with 3'RACE. I would further like to thank Rebecca Bloomer, Julia Qüesta and Peijin Li for useful discussions on *COOLAIR* secondary structure and splicing. Thank you all for being a supportive, inspiring and fun group to be a part of.

I thank all members of the 'Brassica group' (Irwin, Morris and Penfield labs, Martin Trick and Rachel Wells) who have provided scientific and technical advice, useful discussions [and cake]. Thank you to Rachel Wells for sharing *Brassica* plant materials, and to Chantelle Burrowes for all your hard work threshing *Arabidopsis* seeds. Thank you to William Hawkes for assisting with the collection of flowering time data, bagging, and harvesting plants, and for many hours spent threshing seeds, and to Sibyl Batey for giving up your time to help me bag hundreds of plants the day before Christmas Eve 2016.

Most of all, I would like to thank Eleri Tudor, fellow PhD student and friend. Thank you for your support and advice, for sharing plant materials and methods (and photos of Reggie!), for discussions about *FLC* in *Brassica*, and for always being there for a chat.

A special commendation to the support services at the NBI; they have been invaluable during my time here. Thank you to Lionel Perkins and the rest of the Horticultural Services team, who have done a fantastic and diligent job looking after my plants. I would also like to thank the security staff who go above and beyond their duty to make late-night workers feel safe and less alone.

I would like to acknowledge those who have commented on and proofread drafts of this thesis submission. First and foremost, thank you to Judith Irwin for your feedback and comments. Also, thank you to Caroline Dean, Jonathan Cocker and Jeanette Hawkes for your comments and/or proofreading.



Finally, thank you to my family and friends for their love and support over the last four years. Thank you to Jon for being my closest friend and for lifting my spirit. Thank you to my mum, dad, Becki, Will and Annie, without whom I could not be; to my grandads, who were there at the beginning; and to Mystique and Bilbo, for being entirely yourselves. Thank you to each of my friends for making life fun and to Sibyl, Dash and Mabon, who can make even a PhD fun. Lastly, thank you to Grey Cat, Sammy and Eeyore, without whom I would not have got this far.

Contents

Illustrative material

Abbreviations

3S	Shotgun secondary structure determination
A/T/C/G/U	Adenine / Thymine / Cytosine / Guanine / Uracil
AtFLC	*Arabidopsis thaliana FLC*
bp	Base pair(s)
BCS	Bayer CropScience
BnFLC	*Brassica napus FLC*
BoFLC	*Brassica oleracea FLC*
BrFLC	*Brassica rapa FLC*
CBC	Cap-binding protein complex
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIRP	Chromatin isolation by RNA purification
CMCT	1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate
Col-0	Columbia (*A. thaliana* rapid-cycling accession)
COOLAIR	Cold-induced long antisense intragenic RNA
CstF	Cleavage stimulation factor
eQTL	Expression quantitative trait loci
FLC	*Flowering Locus C*
FRI	*FRIGIDA*
GAPDH	*Glyceraldehyde-3-phosphate dehydrogenase*
H3K27me2/3	Histone 3 lysine 27 di/trimethylation
H3K36me3	Histone 3 lysine 36 trimethylation
H3K4me2/3	Histone 3 lysine 4 di/trimethylation
H3K9me2/3	Histone 3 lysine 9 di/trimethylation
HDA	Histone deacetylase
indel	Insertion/deletion
JIC	John Innes Centre
Ka	Thousand years ago
lncRNA	Long non-coding RNA
Lov-1	*A. thaliana* late flowering accession
LSD1	Lysine-specific histone demethylase 1
Ma	Million years ago
miRNA	Micro RNA
mRNA	Messenger RNA
NBI	Norwich BioScience Institutes
ncRNA	Non-protein-coding RNA
nt	Nucleotide(s)
NV	Non-vernalized
PCR	Polymerase chain reaction
PHD	Plant homeodomain protein
poly(A)	Polyadenylation
PPS	PPS02144 (*B. napus* spring cultivar)
PRC2	Polycomb Repressive Complex 2
PSB	Purple Sprouting Broccoli (*B. oleracea* winter cultivar)
PTM	Purple Top Milan (*B. rapa* winter cultivar)
qPCR	Quantitative PCR
QTL	Quantitative trait locus
R018	*B. rapa* spring cultivar
RNA Pol II	RNA Polymerase II
RNAi	RNA interference
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription PCR
r-turn	Right-hand-turn structural motif
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
siRNA	Short interfering RNA
smRNA FISH	Single molecule RNA fluorescence in situ hybridisation
SNP	Single nucleotide polymorphism
tRNA	Transfer RNA
UBC	*Ubiquitin C*
UTR	Untranslated region
V2W/V4W/V6W (etc.)	Two/four/six weeks vernalized (etc.)
Var2-6	*A. thaliana* late flowering accession

1 Introduction

1.1 Overview

This thesis investigates the presence of a group of long non-coding RNAs (lncRNAs), collectively known as *COOLAIR*, in *Brassica* crop species. Their role in vernalization and flowering time is analysed.

COOLAIR lncRNAs were first discovered in the reference plant, *Arabidopsis thaliana*, where they are transcribed in the antisense direction from an important floral repressor, *Flowering Locus C* (*FLC*) (Swiezewski *et al.*, 2009). Protein-coding sense transcripts at *FLC* encode a MADS box transcription factor which quantitatively delays flowering (Michaels and Amasino, 1999). *COOLAIR* initiates at the 3' end of the *FLC* sense transcript and terminates either at a proximal (intron 6) or distal (upstream of 5'UTR) polyadenylation (poly(A)) site (Fig. 1.1A).

COOLAIR plays a role in downregulating levels of the protein-coding sense transcript in the autonomous floral promotion and vernalization pathways, consequently promoting the acceleration of flowering. The ratio of proximal to distal transcripts can influence sense expression. RNA processing reactions that favour the spliced proximal isoform promote FLD-mediated removal of active chromatin marks and, consequently, downregulation of *FLC* in the autonomous flowering pathway (Fig. 1.1B) (Liu *et al.*, 2010; Marquardt *et al.*, 2014). Kinetic coupling of RNA Polymerase II (RNA Pol II) elongation and RNA processing could reinforce this silenced state (Marquardt *et al.*, 2014). Upregulation of *COOLAIR* during vernalization has been linked to the initial downregulation of *FLC* transcription before epigenetic silencing (Swiezewski *et al.*, 2009; Csorba *et al*., 2014).

Arabidopsis and *Brassica* species belong to the same family of flowering plants, the Brassicaceae. Fossil evidence suggests they diverged from a common ancestor around 43 million years ago (Ma) (Beilstein *et al.*, 2010). It is therefore likely that flowering control pathways in *A. thaliana* are at least partially conserved in *Brassica*. *FLC* homologues are wide-spread in flowering plants and functional conservation of the FLC protein has been reported in other members of the Brassicaceae (Kemi *et al.*, 2013; Yi *et al.*, 2014; Taylor *et al.*, 2017). Typically, non-coding RNAs are less well conserved than their protein-coding counterparts (Ponjavic *et al.*, 2007). In 2014, *COOLAIR* antisense transcripts were detected in two *A. thaliana* relatives, *Arabidopsis lyrata* and *Arabis alpina*, and it is suspected that

they share a common function (Castaings *et al.*, 2014). This thesis focuses on *COOLAIR* in *Brassica* species, which include many important oilseed and leafy vegetable crops. It is hoped that analysis of the conservation and behaviour of *COOLAIR* in *Brassica* will shed light on its function (both in *Brassica* and *A. thaliana*). This knowledge could ultimately be used to predict how *COOLAIR* influences variation in flowering time and vernalization response within *Brassica* species.

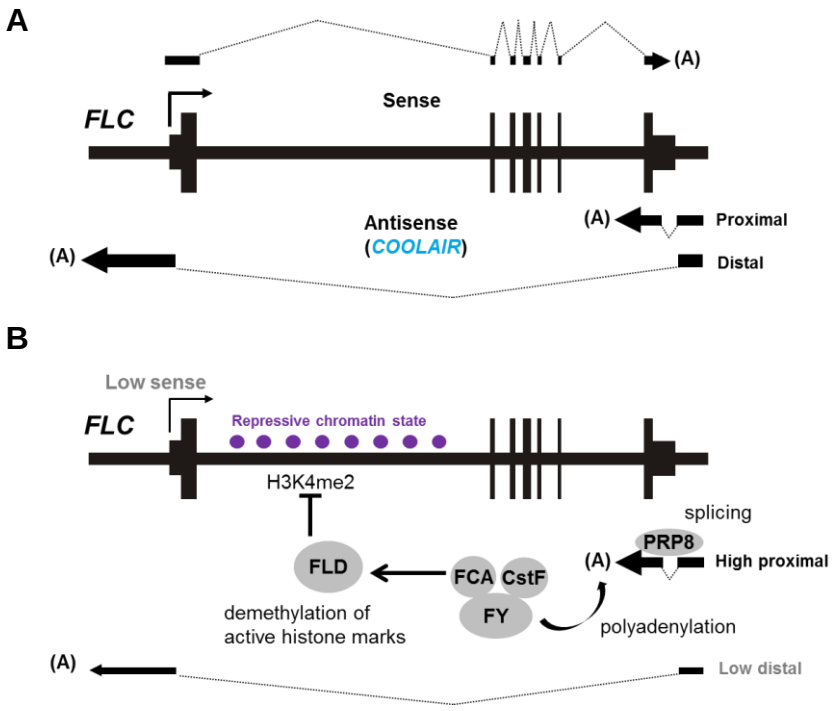


Figure 1.1: Schematic of *FLC* sense and antisense transcripts in *Arabidopsis thaliana*. (A) Schematic of sense and antisense transcripts at the *AtFLC* locus. The shorter antisense transcript is the proximal; the longer, the distal. Black boxes represent exons; dotted lines, introns. Terminal exon arrows indicate the direction of transcription. '(A)' represents the polyadenylated tail. (B) Schematic of FLD-mediated repression of the *FLC* locus in the autonomous flowering pathway described by Liu *et al.* (2010) and Marquardt *et al*. (2014). Briefly, splicing (by PRP8) and polyadenylation (by FCA, FY and CstF) of the proximal *COOLAIR* isoform promotes FLD-mediated demethylation of active H3K4me2 histone marks at *FLC*, pushing it into a repressive state with low protein-coding sense transcription. Grey circles represent autonomous pathway proteins; CstF represents the CstF64 and CstF77 3'end processing factors. Schematics were based on pre-existing versions from the Dean lab.

1.2 The role of *FLC* in flowering time control

1.2.1 Flowering and vernalization

Flowering marks the transition from vegetative to reproductive growth in angiosperms. It is critical to align this transition with environmental conditions that suit seed development and dispersal, i.e. with spring in temperate countries. Consequently, flowering time is tightly regulated by developmental and environmental response pathways. Various environmental signals integrate into a network of floral regulatory genes leading to the final decision on when to flower (Jaeger *et al.*, 2013; Valentim *et al.*, 2015). This network can be divided into at least six major but interconnected genetic pathways: the internal autonomous, age-dependent and gibberellin pathways; and the external photoperiod, thermal responsiveness and vernalization pathways (Fig. 1.2) (Amasino, 2010; Song *et al.*, 2013A; Song *et al.*, 2013B). Through the latter, plants recognise and respond to discrete environmental cues: changes in day length (photoperiod), ambient temperature (thermal responsiveness), and winter cold (vernalization) (Andrés and Coupland, 2012).

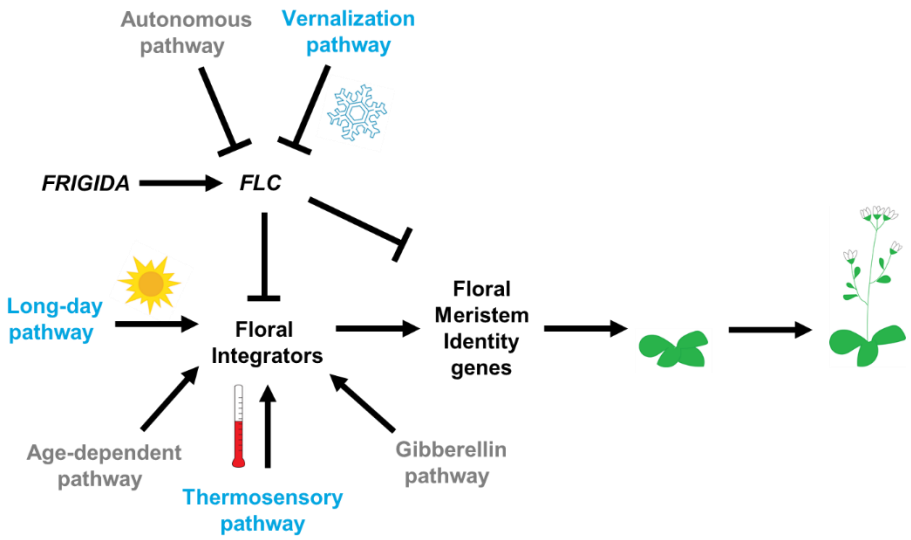


Figure 1.2: A simplified model of flowering time pathways. *FLC* delays flowering by downregulating floral integrators. Endogenous (grey) and exogenous (blue) pathways promote flowering by repressing *FLC* or by directly upregulating floral integrator genes. Floral integrators promote flowering by upregulating floral meristem identity genes, responsible for mediating the transition from vegetative to floral meristem prior to floral organ development. Arrows indicate promotion, flat lines indicate inhibition.

Vernalization describes the acceleration of flowering after winter (or a prolonged period of cold) to align flowering and subsequent seed dispersal with the favourable conditions of spring. Vernalization has been well-characterised in the reference plant *Arabidopsis thaliana* (Song *et al.*, 2013B). It is an example of epigenetic regulation – a stable change in gene expression without change in the underlying DNA sequence. Epigenetic regulation underpins development; genes are switched on or off at different stages, and their new state

maintained through multiple mitotic and sometimes meiotic divisions (Khavari *et al.*, 2010). Examples of transgenerational epigenetic processes in plants include hypermethylation of the *CYCLOIDIA* gene in *Linaria*, which gives rise to radially symmetrical flowers (Cubas *et al*., 1999), hypermethylation of the *Cnr* gene in tomato, which delays ripening (Manning *et al*., 2006), and the widespread loss of repressive histone marks in response to abiotic stresses (Lang-Mladek *et al*., 2010). Cold saturation during vernalization switches the floral repressor *FLC* to an off-state, which is epigenetically maintained through repressive histone modifications (Song *et al.*, 2013B).

1.2.2 Key genes in flowering time control

Floral repressors, activators and other key components of the flowering time control pathways were first identified from early and late *A. thaliana* mutant screens and from naturally occurring accessions (Koornneef *et al.*, 1991; Clarke and Dean, 1993; Koornneef *et al.*, 1994; Lee *et al.*, 1994A; Lee *et al.*, 1994B; Koornneef *et al.*, 1998). Different pathway components converge on the same floral integrator genes that promote floral initiation, including *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*/*AGL20*), *FLOWERING LOCUS T* (*FT*), and *SQUAMOSA BINDING PROTEIN LIKE* (*SPL*) transcription factors. FT is the main component of 'florigen', the supposed flowering hormone common to all flowering plants (Huang *et al.*, 2005). *FLOWERING LOCUS D* (*FD*) encodes a basic leucine zipper domain transcription factor and is involved in FT-mediated floral promotion (Abe *et al.*, 2005).

FT mRNA (and proteins) are produced in the leaf and transported to the shoot apical meristem where FT, together with FD, upregulates *SOC1*, *SPL* transcription factors, and the meristem identity genes, *FRUITFULL* (*FUL*/*AGL8*) and *APETALA 1* (*AP1*) (Andrés and Coupland, 2012). The MADS box transcription factors SOC1 and AGAMOUS-LIKE 24 (AGL24) promote transcription of the meristem identity gene *LEAFY* (*LFY*) (Lee *et al.*, 2008). SPL transcription factors in the shoot apical meristem also promote expression of *SOC1*, *AP1*, *FUL* and *LFY* (Wang *et al.*, 2009A; Yamaguchi *et al.*, 2009). The meristem identity genes *LFY* and *AP1* promote the transition from vegetative to floral meristem at the shoot apex (Weigel *et al.*, 1992; Bowman *et al.*, 1993).

The shoot meristem identity gene *TERMINAL FLOWER 1* (*TFL1*) interacts antagonistically with *LFY* and *AP1* to delay the floral transition, and helps maintain distinct inflorescence shoots post-transition (Bradley *et al.*, 1997). The MADS box transcription factor FLC represses the floral integrators *FT* and *SOC1* (Hepworth *et al.*, 2002; Helliwell

et al., 2006). FLC forms a heterodimer with another MADS box transcription factor, SHORT VEGITATIVE PHASE (SVP), to inhibit flowering (Li *et al.*, 2008). SVP also inhibits the gibberellin (Andrés *et al.*, 2014) and thermosensory (Lee *et al.*, 2007) pathways. *HUA2* is an RNA processing gene which negatively regulates flowering by directly or indirectly enhancing levels of the floral repressors *FLC, SVP, MADS AFECTING FLOWERING 2* (*MAF2*) and *FLOWERING LOCUS M* (*FLM*) (Doyle *et al.*, 2005). *EMBRYONIC FLOWER (EMF) 1* and *2* act in a negative reciprocal regulation loop with floral meristem identity genes to delay flowering until an appropriate developmental stage (Chen *et al.*, 1997).

The autonomous and (after sufficient cold exposure) vernalization pathways indirectly promote floral integrators by downregulating *FLC* (Hepworth *et al.*, 2002; Helliwell *et al.*, 2006). The FCA and FVE autonomous pathway proteins are also involved in the themosensory pathway which represses *FT* until ambient temperature is sufficiently high (Blázquez *et al.*, 2003). *FLC* is upregulated during short periods of extreme low (i.e. pre-winter cold snaps) or high temperatures to prevent premature flowering (Jung *et al.*, 2013; Gan *et al.*, 2014). Homologues of *FLC*, the *MAF 2-5* genes, also play a role in vernalization response, with *MAF2* helping to prevent premature vernalization after a short cold spell (Ratcliffe *et al.*, 2003). In sufficient light (long day) conditions, *FT* is activated by CONSTANS (CO), a zinc finger transcription factor; *CO* is in turn regulated by light-sensing and circadian clock proteins (Putterill *et al.*, 1995; Suárez-López *et al.*, 2001; Tiwari *et al.*, 2010). The latter include the light-sensing circadian clock genes *GIGANTEA* (*GI*) and *FLAVIN KELCH F BOX* (*FKF1*), which release repression of *CO* by CYCLING DOF FACTORs (CDFs) (Fowler *et al.*, 1999; Imaizumi *et al.*, 2003; Sawa *et al.*, 2007).

The endogenous gibberellin floral promotion pathway activates the floral integrators *SOC1* and *FT*; gibberellins also directly upregulate the meristem identity gene *LFY* and are integral for bolting (Wilson *et al.*, 1992; Chandler *et al.*, 1996; Blázquez *et al.*, 1997; Nilsson *et al.*, 1998; Chandler *et al.*, 1999; Moon *et al.*, 2003; Eriksson *et al.*, 2006; for review see: Mutasa-Göttgens *et al.*, 2009). The age-dependent pathway ensures that plants flower only after the juvenile to adult transition. This is at least partially achieved through a micro (mi)RNA based mechanism. In young plants, *SPL* floral integrators are repressed by high levels of the miR156 miRNA; as the plants age, expression levels of miR156 decrease and *SPL* expression increases (Wang *et al.*, 2009A).

Together, these pathways select the best time to flower by balancing developmental and environmental cues.

1.2.3 **FLC** *is an important floral repressor*

A central role for *FLC* as a floral repressor in the autonomous and vernalization pathways became apparent from mutagenesis screens in the 1990s (Koornneef *et al.*, 1994; Lee *et al.*, 1994B). *FLC* was isolated, cloned and characterised by two independent groups (Michaels and Amasino, 1999; Sheldon *et al.*, 1999). Located towards the top of *A. thaliana* chromosome 5, the seven exons of the sense *FLC* transcript encode for a MADS box transcription factor protein, part of a large family of transcription factors which regulate developmental processes (Michaels and Amasino, 1999). Antisense non-protein coding RNAs, *COOLAIR*, are also transcribed from the *FLC* locus (Swiezewski *et al.*, 2009). These initiate at the 3' end of *FLC* and terminate either at a proximal (sense intron 6) or distal (upstream of the sense 5'UTR) site. *COOLAIR* plays a role in downregulating levels of the protein-coding sense transcript in the autonomous and vernalization pathways, described in more detail below.

Overexpression of *FLC*, under the Cauliflower Mosaic Virus (CaMV) 35S strong constitutive promoter, was sufficient to delay flowering in the absence of *FRIGIDA* (*FRI*) (a positive regulator of *FLC*) or autonomous pathway mutants (Michaels and Amasino, 1999; Sheldon *et al.*, 1999). Differential expression of *FLC*, through differences in transgene copy number or genome position, were reflected in the severity of the late flowering phenotype (Michaels and Amasino, 1999; Sheldon *et al.*, 1999). From these experiments, it was suggested that *FLC* repression is quantitative; it acts via a rheostat mechanism whereby the timing of flowering reflects *FLC* dosage (Michaels and Amasino, 1999; Sheldon *et al.*, 1999).

FLC acts as a floral repressor by directly downregulating the floral integrators *SOC1* and *FT* (Lee *et al.*, 2000; Samach *et al.*, 2000; Michaels and Amasino, 2001; Hepworth *et al.*, 2002). FLC binds *SOC1* and *FT* chromatin *in vivo* as part of a multimeric protein complex (Helliwell *et al.*, 2006). This repressive complex is thought to include the SVP transcription factor, as FLC and SVP regulate *SOC1* and *FT* in a co-dependent manner (Li *et al.*, 2008). FLC may also interact with chromatin-remodelling proteins to induce a stable repressive state at *SOC1* and *FT* (Helliwell *et al.*, 2006).

1.2.4 FRIGIDA *upregulates* FLC

Early research revealed that *FRIGIDA* (*FRI*) was responsible for late flowering phenotypes in Stockholm, Sf-2, Co-4, Ge-2 and Zu-0 *Arabidopsis* accessions, whilst non-functional *FRI* alleles were responsible for the early flowering phenotype in Landsberg *erecta* and Columbia (Col-0) (Sanda *et al.*, 1997; Johanson *et al.*, 2000). *FRI* maps to the top of *A. thaliana* chromosome 4 and encodes a 609 amino acid protein with two coiled-coil domains (Sanda *et al.*, 1997; Johanson *et al.*, 2000).

An interaction with *FLC* was first indicated by the synergistic nature of the late-flowering Sf-2 *FRI* and Sf-2 *FLC* alleles in a Landsberg *erecta* background (Koornneef *et al.*, 1994). When crossed into an *flc-3* null background the *FRI* late flowering phenotype was eliminated (Michaels and Amasino, 2001). Thus, *FRI* only delays flowering in the presence of an active *FLC*. Expression levels of *FLC* and *FRI* were found to correlate: with late flowering functional *FRI* alleles, *FLC* expression was high; with early flowering non-functional *FRI* alleles, expression was low (Sheldon *et al.*, 1999). Northern blot analysis revealed that *FLC* transcript levels increase in the presence of *FRI* (Michaels and Amasino, 1999). Consequently, the role of *FRI* as transcriptional activator of *FLC* was proposed (Michaels and Amasino, 1999; Sheldon *et al.*, 1999).

Several mutants were identified which suppress *FRI* activity without suppressing autonomous pathway mutants, implying a specific interaction. These included: *FRIGIDA-LIKE* (*FRL*) *1* and *2* (Michaels *et al.*, 2004), zinc finger domain genes, including *SUPPRESSOR OF FRIGIDA 4* (*SUF4*) (Kim *et al.*, 2006; Kim and Michaels, 2006) and *FRIGIDA-ESSENTIAL 1* (*FES1*) (Schmitz *et al.*, 2005), and putative leucine zipper domain protein-encoding genes *FLC EXPRESSOR* (*FLX*) (Andersson *et al.*, 2008) and *FLC EXPRESSOR LIKE 4* (*FLL4*) (Lee and Amasino, 2013). It was hypothesised that these specific regulators form a protein complex which transcriptionally upregulates *FLC* (Choi *et al.*, 2011). Overexpression of any one proposed element of this complex – *FRI*, *SUF4*, *FLX*, *FRL1*, *FRL2* and *FES1* – cannot supress a mutation in any other (Choi *et al.*, 2011). This supports a pathway of *FLC* regulation involving all six genes. Possibly, FRI acts as a scaffold, interacting with the other proteins to form a complex and binding *FLC* via SUF4 (Kim *et al.*, 2006; Choi *et al.*, 2011). Transcriptional activation of *FLC* could be promoted by the transcriptional activators FES1 and FLX, and by recruitment of general transcription factors, chromatin remodellers (the SWR1-C and PAF1-like complexes), and RNA Pol II (He *et al.*, 2004; Kim and Michaels, 2006; Choi *et al.*, 2011). Upregulation of *FLC* by *FRI* is accompanied by enrichment of H3K4me3 (histone 3 lysine 4 trimethylation) and

H3K36me3 (histone 3 lysine 36 trimethylation) active histone marks across the *FLC* gene body (He *et al.*, 2004).

FRI additionally interacts with a nuclear cap-binding protein complex (CBC) (Bezerra *et al.*, 2004; Geraldo *et al.*, 2009). Mutations in the CBP80 and CBP20 subunits of this complex supressed high *FLC* levels in both *FRI+* and autonomous pathway mutant backgrounds (Bezerra *et al.*, 2004; Geraldo *et al.*, 2009). A more specific function in the *FRI* pathway is implicated because CBP80 specifically interacts with FRI, and not autonomous or vernalization pathway components (Geraldo *et al.*, 2009). The nuclear CBC binds to the 5' cap of mRNAs and is involved in RNA processing, stability and export. FRI-CBC likely upregulates *FLC* by directly increasing the proportion of mRNAs with a 5' cap.

1.2.5 The autonomous pathway downregulates FLC

Recessive mutants *fca*, *fld*, *fve*, *fpa*, *flk*, *ld*, and *fy* delay flowering but are still responsive to environmental cues; their late flowering phenotypes are reversed after cold treatment or in long day conditions (Koornneef *et al.*, 1991; Lee *et al.*, 1994). This is because they disrupt the autonomous floral promotion pathway. Autonomous pathway proteins promote flowering by reducing *FLC* expression levels (Lee *et al.*, 1994B; Sanda and Amasino, 1996; Sheldon *et al.*, 1999; Sheldon *et al.*, 2000; Michaels and Amasino, 2001).

The relationships between members of this group are complex, with double mutant analysis revealing additive and synergistic, as well as epistatic, interactions (Koornneef *et al.*, 1998). This can be explained by devising functional sub-categories within the group. FVE and FLD are chromatin modifiers, interacting with target loci independently or with other autonomous pathway members (Ausín *et al.*, 2004; Lim *et al.*, 2004; Bäurle and Dean, 2008). FVE has been reported to promote histone deacetylation at *FLC* through association with HISTONE DEACETYLASE 6 (HDAC6) (Gu *et al.*, 2011). It also directly antagonizes upregulation of *FLC* by the *FRI*-complex (Lee and Amasino, 2013). *FLD*, an orthologue of the human Lysine-Specific Demethylase 1 (*LSD1*; see Shi *et al.*, 2004) chromatin remodelling protein, promotes histone demethylation at *FLC* in an *FCA* or *FPA* dependent manner (He *et al.*, 2003; Liu *et al.*, 2007). *FCA* (Macknight *et al.*, 2002), *FPA* (Schomburg *et al.*, 2001), and the later identified *FLK* (Lim *et al.*, 2004) encode nuclear RNA-binding proteins, with RRM, RRM, and KH motifs respectively. *FY* encodes a RNA 3' processing protein, which interacts with FCA to promote the use of alternate 3' polyadenylation sites (Simpson *et al.*, 2003). FCA also interacts with FLD to promote

chromatin modifications at a target site (Liu *et al.*, 2007). FPA contains three RNA binding domains and interacts with FLD independently of FCA and FY (Schomburg *et al.*, 2001; Bäurle and Dean, 2008; Hornyik *et al.*, 2010; Liu *et al.*, 2010). *LD* has been mapped to *A. thaliana* chromosome 4 and encodes a putative transcription factor (Lee *et al.*, 1994A; Auckerman *et al.*, 1999). The autonomous pathway proteins regulate their targets in a post-transcriptional manner via chromatin modifications or RNA processing, but do not all act in the same molecular mechanism (Bäurle and Dean, 2008). They have multiple roles across the *A. thaliana* transcriptome, but *FLC* is thought to be a particularly sensitive target (Bäurle and Dean, 2008; Sonmez *et al.*, 2011).

FCA encodes a protein with a WW protein-protein interaction domain and two RNA-binding domains (Macknight *et al.*, 1992). It has two unusual traits: it is translated exclusively from a noncanonical CUG rather than AUG translation start site (Simpson *et al.*, 2010) and it has four splice variants, with only one pertaining to the functional protein (Macknight *et al.*, 1997). The functional full-length transcript is normally expressed at 35% of total levels, which limits FCA protein availability (Macknight *et al.*, 2002). Overexpression of *FCA* does not increase levels of the functional transcript because at a certain threshold the use of an alternate poly(A) site within intron three, and consequently production of a non-functional transcript, is promoted (Queseda *et al.*, 2003). The FY RNA 3' processing protein was found to interact with FCA via its WW domain, and to promote the selection of this proximal poly(A) site (Simpson *et al.*, 2003). Thus, auto-regulation of *FCA* limits flowering at different developmental time-points (Queseda *et al.*, 2003).

The discovery of *FCA* auto-regulation led to the hypothesis that FCA and FY might downregulate *FLC* by a similar post-transcriptional mechanism (Queseda *et al.*, 2003). FCA binds *FLC* chromatin directly at exon 6/intron 6 (Sheldon *et al.*, 2002; Liu *et al.*, 2010). Surprisingly this region did not correspond to an *FLC* sense polyadenylation site; however, it did correspond to the proximal *COOLAIR* antisense polyadenylation site (Swiezewski *et al.*, 2009; Liu *et al.*, 2010). The ratio of the proximal and distal classes (Class I and Class II *COOLAIR* transcripts, respectively) were considered significant, because Class II distal transcripts were expressed at a higher level relative to Class I proximal transcripts in *fca-9, fy* and *fld-3* mutants, whilst the opposite pattern was seen in wildtype plants (Liu *et al.*, 2010). This suggested that 3' processing by FCA and FY targets *FLC* antisense, rather than sense, transcripts, promoting the use of the proximal polyadenylation site (Liu *et al.*, 2010). The same low ratio of proximal:distal transcripts was noted in *CstF64* and *CstF77 suppressor of overexpressed fca* (*sof*) mutants,

components of a suspected RNA 3' processing complex (Liu *et al.*, 2010). It was therefore proposed that FCA interacts with FY to target CstF-dependent 3' processing to the proximal *FLC* antisense polyadenylation site (Liu *et al.*, 2010). How upregulation of the proximally polyadenylated transcript contributes to downregulation of the sense transcript is not fully understood.

The current model for autonomous pathway regulation of *FLC* is presented in Fig. 1.1B, and briefly as follows. The RNA recognition proteins FPA and FCA interact (independently) with the poly(A) specificity factor FY and the cleavage stimulation factors CstF64 and CstF77 to promote use of the proximal poly(A) site (Hornyik *et al.*, 2010; Liu *et al.*, 2010). Use of this site and efficient processing of the proximal transcript are further promoted by the spliceosome component PRP8 and the Positive Transcription Elongation Factor b (PTEFb) component, CYCLIN DEPENDENT KINASE GROUP C2 (CDKC;2) (Marquardt *et al.*, 2014; Wang *et al.*, 2014A). Splicing and polyadenylation of the proximal transcript decreases levels of H3K4me2/me3 active chromatin marks across the *FLC* gene body, through promotion of the FLD H3K4 demethylase (Liu *et al.*, 2007). In turn, the FLD-mediated repressive chromatin state promotes use of the proximal antisense poly(A) site, potentially through a cotranscriptional coupling mechanism (for review, see: Bentley, 2014), with slower RNA Pol II elongation leading to early transcript termination (Marquardt *et al.*, 2014). These processes help to stabilise *FLC* in an inactive state, reducing both sense and antisense transcription.

Interestingly, small 24-30 nt antisense RNAs have also been detected at the 300 bp *COOLAIR* promoter (Swiezewski *et al.*, 2007). Mutant analysis indicated that the 24 nt RNAs (but not the 30 nt RNAs) are processed by standard plant RNAi machinery, including DICER-like (DCL) 3, RNA-dependent RNA Pol II and RNA Pol IV (Swiezewski *et al.*, 2007). Mutants defective in small RNA production had higher *FLC* sense expression. The 300 bp promoter region is enriched in H3K9me2; the small RNAs could recruit chromatin remodelling proteins, like FLD, to this region and so disrupt *FLC* expression, perhaps via inefficient 3' transcriptional termination (Liu *et al.*, 2007; Swiezewski *et al.*, 2007). Recent focus has been on the lncRNAs that are transcribed from the *COOLAIR* promoter (likely the small RNA precursors), but a role for the small RNAs themselves has not been ruled out.

1.2.6 The vernalization pathway epigenetically silences **FLC**

FLC *is downregulated in the cold*

Balance between autonomous pathway downregulation and *FRIGIDA* upregulation maintains *FLC* at a level sufficient to delay flowering. This can be overcome by prolonged cold – vernalization (Koornneef *et al.*, 1994; Michaels and Amasino, 1999; Sheldon *et al.*, 1999). *FLC* expression is downregulated quantitatively in the cold; different lengths of cold treatment are required to saturate the vernalization requirement of different *Arabidopsis* accessions (Michaels and Amasino, 1999; Sheldon *et al.*, 1999; Sheldon *et al.*, 2000). Initial levels of *FLC* expression and the stability of *FLC* repression are thought to determine the length of cold required to saturate vernalization requirement; rate of downregulation is similar across a wide range of ecotypes (Shindo *et al.*, 2006).

The silenced *FLC* state is maintained post-vernalization after return to warm (Michaels and Amasino, 1999). *FLC* repression is therefore mitotically stable. Plant cells remember that winter (vernalization) has passed, even if flowering does not occur until many months after (i.e. when day length and ambient temperature are sufficient). *FLC* levels are re-set in the next generation, so are not meiotically stable (Michaels and Amasino, 1999; Sheldon *et al.*, 2000). Presumably, the changes induced in vernalization to stabilise *FLC* are lost during meiosis. This is specific to sexual reproduction because transcriptional repression at *FLC* is maintained after *in vitro* regeneration (Nakamura and Hennig, 2017). *FLC* repression is maintained through multiple mitotic cycles and is therefore epigenetic in nature (Gendall *et al.*, 2001). Early work suggested a link with DNA methylation (Burn *et al.*, 1993; Sheldon *et al.*,1999). However, when bisulphite sequencing revealed unchanged DNA methylation levels pre- and post-vernalization at the *FLC* locus (Bastow *et al.*, 2004), focus turned to histone modifications.

Chromatin immunoprecipitation (ChIP) experiments revealed differences in histone marks pre- and post-vernalization at *FLC* (Bastow *et al.*, 2004). The *FLC* locus is enriched in H3K27me2/3 and H3K9me2 repressive histone marks post-vernalization, replacing the H3K4me2/3 and H3K36me3 active histone marks found pre-vernalization (Bastow *et al.*, 2004; Sung and Amasino, 2004; Schubert *et al.*, 2006; Sung *et al.*, 2006). These changes are accompanied by vernalization-induced histone deacetylation (Sung and Amasino, 2004). A 2.8 kb region within intron 1, which had previously been found essential for the maintenance of *FLC* repression after cold (Sheldon *et al.*, 2002), contained higher levels of the repressive histone marks relative to surrounding areas (Bastow *et al.*, 2004). This region was later termed the ‘nucleation region’ because it is where H3K27me3

predominantly increases during cold, before spreading across the whole *FLC* locus post-cold (De Lucia *et al.*, 2008). H3K27 trimethylation does not, however, spread to neighbouring genes, despite having the ability to do so in a transgenic context (Sheldon *et al.*, 1999; Finnegan *et al.*, 2004; Schubert *et al.*, 2006). This suggests there are *cis* factors up- or downstream of the coding region which halt the spread of heterochromatin.

Key players in FLC *silencing*

The late flowering *vrn1, vrn2* and *vrn5 Arabidopsis* lines contain mutations in genes involved in vernalization perception and response. These mutations affect vernalization response, but do not alter non-vernalized flowering time [*VRN2* is an exception to this; it may be involved in an unrelated floral promotion pathway before vernalization] (Chandler *et al.*, 1996; Sheldon *et al.*, 1999; Sheldon *et al.*, 2000; Gendall *et al.*, 2001). Accordingly, reduction of the *FLC* transcript during vernalization was slower in Landsberg *erecta vrn1 fca-1* and *vrn2 fca-1* double mutants than in the *fca-1* single mutant; the Landsberg *erecta fca-1* background was selected for these experiments as it is late flowering and responds strongly to vernalization (Sheldon *et al.*, 2000). Significantly, *FLC* mRNA levels in *vrn1* and *vrn2* decrease but are not maintained at a low level after return to the warm (Levy *et al.*, 2002). The vernalization-induced changes in H3K4me2 were lost in the *vrn1* mutant, H3K27me2 in *vrn2*, and H3K9me2 in both (Bastow *et al.*, 2004; Sung and Amasino, 2004). This suggested that the VRN1 and VRN2 proteins epigenetically silence *FLC* by establishing and/or maintaining histone modifications associated with a repressive state.

VRN2 encodes a zinc finger domain nuclear protein, homologous to the *Drosophila* Polycomb-group protein and Polycomb Repressive Complex 2 (PRC2) component, Supressor of zeste 12 (Su(z)12) (Gendall *et al.*, 2001). PRC2 in *Drosophila* and mammals maintains repressed chromatin states through H3K27 trimethylation of histone tails (Kuzmichev *et al.*, 2002). The VRN2 Su(z)12 homologue was later confirmed to be a core component of the plant PRC2 complex, alongside proteins encoded by *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*), *CURLY LEAF* (*CLF*), *SWINGER* (*SWN*) and *MSI1* (Wood *et al.*, 2006; De Lucia *et al.*, 2008). The PRC2 complex was found to interact with the VERNALIZATION INSENSITIVE 3 (VIN3) plant homeodomain (PHD) protein during vernalization (Wood *et al.*, 2006). *VRN1* encodes a putative B3 domain nuclear protein which can bind non-sequence specific DNA, including *FLC, in vitro* (Levy *et al.*, 2002). VRN1 may be involved in directing H3K9 methyltransferases to the *FLC* locus via a PRC1-like mechanism (Mylne *et al.*, 2006). Interaction between VRN1 and VRN2 was

ruled unlikely by yeast two-hybrid assays, thus supporting two independent mechanisms of *FLC* repression (Levy *et al.*, 2002).

VIN3 expression is induced by (and only expressed during) cold treatment, and therefore acts as a measure of cold length (Sung and Amasino, 2004). The *vernalization responsive 3 (vin3)* mutant completely blocked vernalization response in a Col *FRI* background (Sung and Amasino, 2004). In contrast to the *vrn* mutants there was no initial reduction in *FLC* expression during the cold (Sung and Amasino, 2004). VIN3 is part of a family of PHD proteins, with homologues in *A. thaliana* including VRN5, VEL1, VEL2 and VEL3 (Greb *et al.*, 2007). VRN5 and VIN3 interact via their shared C terminal domain to form a heterodimer (Greb *et al.*, 2007). *vin3* and *vrn5* mutants both have reduced H3 deacetylation and H3K27 trimethylation at *FLC* after cold treatment (Sung and Amasino, 2004; Greb *et al.*, 2007). The PHD proteins VRN5, VIN3 and VEL1 were found to form a complex with PRC2 at the *FLC* nucleation region during the cold (De Lucia *et al.*, 2008). This is facilitated by specific binding of the B3 domains of a VIVIPAROUS1/ABI3-LIKE factor 1 (VAL1) homo- or hetero-dimer to RY motifs in the *FLC* nucleation region (Qüesta *et al.*, 2016). VAL1 recruits the ASAP (apoptosis- and splicing-associated protein) complex and HISTONE DEACETYLASE 19 (HDA19), which help to shut-down transcription and reduce histone acetylation at *FLC*, thus allowing the PHD-PRC2 complex to nucleate (Qüesta *et al.*, 2016). After cold, the PHD-PRC2 complex spreads across the whole *FLC* locus, promoting H3K27 trimethylation.

LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), the *Arabidopsis* homologue of the heterochromatin-associated HP1 protein, also plays a role in the maintenance of *FLC* repression during cold (Mylne *et al.*, 2006; Sung *et al.*, 2006). *FLC* transcript levels returned to non-vernalized levels after cold treatment in *lhp1* mutants (Mylne *et al.*, 2006; Sung *et al.*, 2006). In wildtype plants, enriched levels of LHP1 at the *FLC* locus corresponded with enriched levels of H2K9me2 during vernalization, whilst in the *lhp-1* mutant H3K9me2 was initially present but not maintained post-cold (Sung *et al.*, 2006). A role in maintenance, but not initiation, of H3K9me2 is supported (Sung *et al.*, 2006). A model was proposed in which LHP1 binds VRN1-dependent H3K9me2 and induces heterochromatin formation at *FLC* (Mylne *et al.*, 2006; Sung *et al.*, 2006). *In vivo*, however, LHP1 was found to exclusively bind H3K27me3 chromatin marks and is therefore likely to be involved in their maintenance at *FLC* after cold (Turck *et al.*, 2007).

Long non-coding transcripts are upregulated during vernalization

The proximal and distal antisense transcripts, which regulate *FLC* in the autonomous pathway, are upregulated in the cold (Swiezewski *et al.*, 2009). These transcripts were named *COOLAIR* (cold induced long antisense intragenic RNA) as a parallel to the mammalian lncRNA *HOTAIR*. *COOLAIR* upregulation was detected by 14 days' cold exposure; this is before maximal VIN3 accumulation and therefore likely precedes PHD-PRC2 complex formation (Swiezewski *et al.*, 2009). *COOLAIR* may instead be involved in initial *FLC* downregulation at a transcriptional level. When the *COOLAIR* promoter was fused to the 3' end of a 35S:*GFP* construct, the induction of *GFP* antisense transcripts in the cold correlated with construct silencing (Swiezewski *et al.*, 2009). From this, the authors suggested that the act of transcription, rather than specific nucleotide sequences within *COOLAIR*, was sufficient for sense regulation. It was later found that *COOLAIR* physically associates with *FLC* chromatin and mediates reduction of H3K36me3 in the nucleation region during cold (Csorba *et al.*, 2014). Paradoxically, this suggests *COOLAIR* transcripts themselves may be important for *FLC* downregulation, perhaps by recruiting a H3K36 demethylase or by inhibiting a methyltransferase (Csorba *et al.*, 2014). T-DNA insertions in the promoter region of *COOLAIR* in a Col *FRI* background did not impact *FLC* repression during the cold (Helliwell *et al.*, 2011). This led to speculation that *COOLAIR* was not important for the vernalization pathway. It is possible that *COOLAIR* was not fully deactivated in the T-DNA lines or that it acts redundantly with other processes.

Another cold-induced lncRNA, *COLDAIR*, is transcribed in the sense direction from a cryptic promoter within the previously discovered Vernalization Response Element (VRE) in *FLC* intron 1, critical for maintaining *FLC* silencing post-vernalization (Sheldon *et al.*, 2002; Sung *et al.*, 2006; Heo and Sung, 2011). Maximum expression of *COLDAIR* is later than *COOLAIR*, at around 20 days after cold exposure (Heo and Sung, 2011). PRC2 components were reported to bind *COLDAIR in vitro* and transiently *in vivo* during cold (Heo and Sung, 2011; Kim and Sung, 2017). It was therefore suggested that *COLDAIR* acts as a scaffold, directing the PHD-PRC2 complex and other chromatin modifiers to the *FLC* locus (Heo and Sung, 2011). This mechanism is similar to that of the mammalian lncRNA, *HOTAIR*, which apparently acts as a scaffold to direct PRC2 to metastasis suppressor genes (Gupta *et al.*, 2010). PRC2 has been found to promiscuously bind multiple different RNAs *in vitro* and has a higher affinity for long transcripts; therefore, supposedly specific interactions between lncRNAs and PRC2 components should be treated with caution (Davidovich *et al*., 2013).

RNAi knock-down of *COLDAIR* compromised vernalization response in many (but not all) transformant lines; a late flowering phenotype was consistent with lack of stable *FLC* repression in two lines that were tested (Heo and Sung, 2011). Vernalization response was not, however, affected in the *COLDAIR* RNAi lines after treatment with longer periods of cold, suggesting redundancy with other vernalization-mediated *FLC*-independent pathways or compensatory action by other components of this pathway (Heo and Sung, 2011). *COLDAIR* is 5' capped but not polyadenylated. This can make detection difficult; it was not successfully found in *A. thaliana* by the Dean lab (personal communication) or in *A. thaliana* or *A. alpina* by Castaings *et al.* (2014). Another sense lncRNA transcribed from the 5' end of *FLC*, *COLDWRAP*, was recently detected (Kim and Sung, 2017; Zhe Wu, Dean lab, unpublished). *COLDWRAP* and *COLDAIR* are thought to promote a repressive chromatin loop at the 5' end of *FLC* (Kim and Sung, 2017).

Current model for the epigenetic silencing of FLC

Changes in *FLC*, *COOLAIR* and *VIN3* gene expression during the cold are summarised in Fig. 1.3.

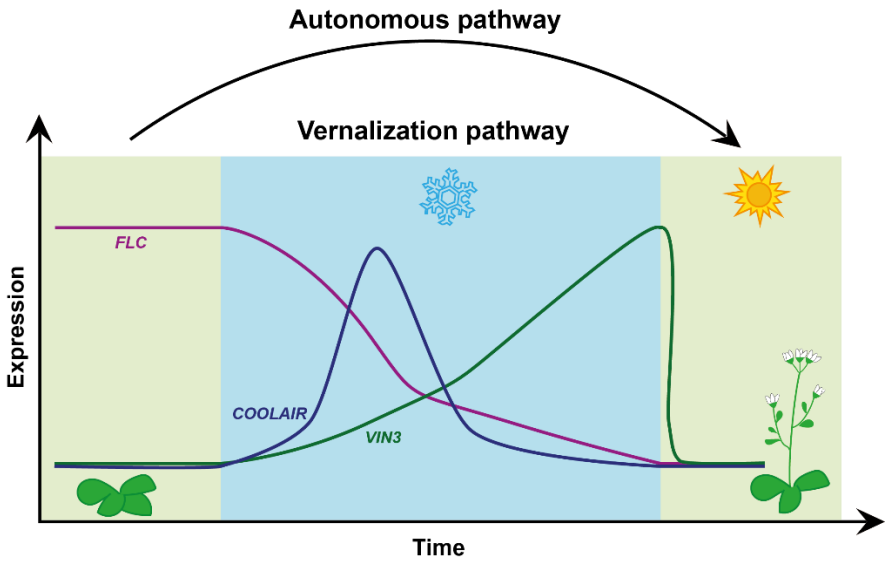


Figure 1.3: Changes in gene expression during vernalization. *COOLAIR* is rapidly upregulated in the cold, peaking at around 14 days, and contributing to initial downregulation of *FLC* sense transcription. *VIN3* is quantitatively upregulated during cold and is involved in the epigenetic silencing of the *FLC* locus.

COOLAIR antisense transcripts drive rapid reduction in *FLC* sense expression levels after transfer to cold (Swiezewski *et al.*, 2009). Modelling suggests that the chromatin state at *FLC* is highly dynamic; switching between chromatin marks gives rise to bistable epigenetic states that are reinforced by positive feedback loops (Angel *et al.*, 2011). Active H3K36me3 and repressive H3K27me3 chromatin marks at *FLC* exhibit opposing patterns; they are rarely found together on the same histone tail (Yang *et al.*, 2014). *COOLAIR*

mediates removal of H3K36me3 at *FLC* in the cold and is therefore thought to be involved in the gradual switch from active to repressive chromatin state during vernalization (Csorba *et al.*, 2014; Yang *et al.*, 2014). The physical interaction of the 5' and 3' ends of *FLC* in a gene loop (thought to influence sense transcription dynamics) is also disrupted in the first weeks of cold (Crevillén *et al.*, 2013). Formation of a repressive chromatin loop between the 5' end of *FLC* and the 3' end of the first intron may help to prevent this from reforming (Kim and Sung, 2017).

The PRC2 complex binds the whole length of the *FLC* locus pre-vernalization (De Lucia *et al.*, 2008). The PHD protein VIN3 is upregulated during cold and associates with the constitutively expressed VRN5 and VEL1 PHD proteins and PRC2 (Sung and Amasino, 2004; Greb *et al.*, 2007; De Lucia *et al.*, 2008). VAL1 binds *FLC* at the RY motifs in the nucleation region and recruits the ASAP complex and HDA19 to shut-down transcription (Qüesta *et al.*, 2016). This allows the PHD-PRC2 complex to associate with the nucleation region, promoting the addition of H3K27me3 during cold (De Lucia *et al.*, 2008; Yang *et al.*, 2014). The level of H3K27me3 that accumulates in this region quantitatively reflects the duration of the cold period (Angel *et al.*, 2011). Eventually, a saturation point is reached and/or the plant is returned to ambient temperatures (Angel *et al.*, 2011). In the warm, if length of cold was sufficient for saturation, then the PHD-PRC2 complex spreads across the whole locus, accompanied by spread of H3K27me3 (De Lucia *et al.*, 2008). LHP1 helps to maintain H3K27me3 after return to the warm (Turck *et al.*, 2007). VRN1 is thought to promote the repressive H3K9me3 mark, and may have a role in maintenance of the silenced state because it stays associated with *FLC* during mitosis (Mylne *et al.*, 2006). Thus, a stable repressive chromatin state exists and *FLC* is epigenetically silenced.

The accumulation of H3K27me3 in the nucleation region quantitatively reflects length of cold; this is because it is promoted by cold-induced VIN3 association with the other PHD proteins and PRC2 (Angel *et al.*, 2011). It was proposed that the extent of H3K27me3 in the nucleation region (i.e. the duration of cold) determines whether *FLC* loci in a single cell switch to a repressive state (Angel *et al.*, 2011). If H3K27me3 does not spread across the locus on return to the warm then *FLC* expression returns to pre-vernalized levels and the cell remains active. Cells are therefore found either in a repressive off-state or an active on-state after cold and actual flowering time is determined by the ratio of on:off cells (Angel *et al.*, 2011). 'Digital' memory, with cells being in either an on- or off-state, rather than 'analog' memory where they are continuously varying, was confirmed experimentally (Angel *et al.*, 2015). The memory of cold is stored locally in the *FLC* chromatin in *cis*;

hence one copy of *FLC* within a cell can be on, and the other off (Berry *et al.*, 2015). This memory can then be passed onto daughter cells after mitotic division (Berry *et al.*, 2015). This explains why the stability of *FLC* repression, via the accumulation of epigenetic modifications, is more important than rate or degree of downregulation in determining saturation time (Shindo *et al.*, 2006).

1.2.7 Natural variation in Arabidopsis *flowering times*

Natural accessions of *Arabidopsis thaliana* are found across temperature zones of the Northern hemisphere (Alonso-Blanco *et al.*, 2009). There is vast natural variation in the flowering times and vernalization responses of these accessions (Li *et al.*, 2014; Duncan *et al.*, 2015). This reflects strong selective pressure to adapt flowering time to a new environment. Adaptations are complex and can reflect large and small-scale environmental change, with a latitudinal cline reported in some studies (Caicedo *et al*., 2004; Stinchcombe *et al.*, 2004), and a microclimatic or local response in others (Johanson *et al.*, 2000; Shindo *et al.*, 2006; Ågren *et al*., 2017). Natural variation in optimal vernalization temperatures has been reported across habitats with varying winter temperatures (Wollenberg and Amasino, 2012; Duncan *et al.*, 2015). Effective vernalization temperature was surprisingly high in some of the more Northern (cold-latitude) accessions; this allows plants to vernalize before snowfall in autumn (Duncan *et al.*, 2015). Hence, some accessions complete their life-cycle rapidly; whilst others remain vegetative over winter and flower in the favourable conditions of spring. Natural variation between late and early flowering *A. thaliana* accessions has been mapped to a few major-effect QTLs, which include *FLM*, *HUA2*, *VIN3*, *FRI* and *FLC* (Werner *et al.*, 2005A and 2005B; Shindo *et al.*, 2006; Strange *et al.*, 2011; Sánchez-Bermejo *et al*., 2012; Ågren *et al*., 2017).

Vernalization requirement describes whether a plant requires vernalization to flower or not. It has been selected for in crops in the form of spring (no vernalization requirement) versus winter (vernalization requirement) cultivars, but is also found naturally in *A. thaliana* accessions. Vernalization requirement has frequently been mapped to *FRI* (Johanson *et al.*, 2000; Strange *et al.*, 2011; Sánchez-Bermejo *et al*., 2012). Assuming a functional *FLC* allele, an ecotype with an active *FRI* normally requires vernalization to flower. A study of 38 differentially flowering ecotypes revealed at least five groups of non-functional *FRI* alleles that have evolved independently over time to confer early flowering phenotypes (Johanson *et al.*, 2000). Further studies supported the correlation between early flowering accessions and independently evolved loss-of-function mutations in *FRI* (Le Corre *et al.*, 2002; Gazzani *et al.*, 2003; Shindo *et al.*, 2005; Werner *et al.*, 2005A). This suggests a

strong recent selection pressure for early flowering ecotypes, perhaps linked to warmer interglacial conditions after the Last Glacial Maximum ~26 thousand years ago (Ka) (Le Corre *et al.*, 2002). Loss-of-function *FLC* alleles have also been identified; these eliminate vernalization requirement, but are less common in naturally occurring populations (Johanson *et al.*, 2000; Werner *et al.*, 2005A). Weak *FLC* alleles with low sense expression, however, can confer a rapid-cycling spring life habit and have arisen independently at least twice over evolutionary time (in the Shakhdara and Landsberg *erecta* accessions) (Michaels *et al.*, 2003); whereas the Lz-0 accession with no functional *FRI* but high *FLC* expression behaves as a winter annual (Werner *et al.*, 2005A).

Plants with vernalization requirement still exhibit a wide range of flowering times and this is largely due to epigenetic differences in their vernalization response (Shindo *et al.*, 2006). Correlation was found between vernalization response and initial *FLC* expression levels (Sheldon *et al.*, 2000), but this relationship becomes more variable in a non-functional *FRI* background (Shindo *et al.*, 2005). Different lengths of cold are required to saturate the vernalization response of different accessions, and this was associated with stability of *FLC* repression, rather than the rate or extent of downregulation (Shindo *et al.*, 2005; Shindo *et al.*, 2006).

cis variation at *FLC* can account for differences in flowering time and vernalization response (Coustham *et al.*, 2012; Li *et al.*, 2014; Li *et al.*, 2015; Sánchez-Bermejo *et al.*, 2012; Méndez-Vigo *et al.*, 2016; Qüesta *et al.*, 2016). Variation in the coding sequence can be limited; for example, there is 100% amino acid identity between the active Col-0 and C24 and the weak Landsberg *erecta FLC* alleles (Sheldon *et al.*, 2000). More frequently, polymorphisms in non-coding regions have been reported to affect flowering times. One example is the long (nine week) vernalization requirement of the *A. thaliana* Lov-1 ecotype [compared with the four-week Col *FRI* vernalization requirement], which was mapped to *cis* variation in non-coding regions of *FLC* (Coustham *et al.*, 2012). The sequence polymorphisms in Lov-1 were linked to reduction in H3K27me3 before and after vernalization (Coustham *et al.*, 2012). A 50 bp deletion in the 5'UTR of *FLC* was associated with rapid vernalization in the late flowering Ll-0 accession (Sánchez-Bermejo *et al.*, 2012). Of particular relevance [to this thesis] is the non-coding polymorphism that drives late flowering in the Var2-6 accession. The causative SNP is adjacent to a distal *COOLAIR* splice site; it alters distal splicing and consequently impedes *COOLAIR* regulation of *FLC*, giving rise to high sense expression (Li *et al.*, 2015).

1.3 Mechanisms for *COOLAIR* regulation of *FLC*

1.3.1 Long non-coding RNAs

Protein-coding mRNAs account for less than 2% of total genomic transcripts in humans (Kim and Sung, 2012). The remainder, the non-protein coding (nc)RNAs, can originate from both coding and intergenic regions and can be broadly divided into two categories: infrastructural and regulatory (Kim and Sung, 2012). Infrastructural ncRNAs are involved in the splicing, transcription and translation of genes i.e. the ribosomal (r), transfer (t), small nuclear (sn) and small nucleolar (sno)RNAs. Regulatory ncRNAs are further removed from protein synthesis pathways, instead serving as transcriptional or post-transcriptional regulators of gene expression (Turner and Morris, 2010).

The best studied regulatory ncRNAs are the small, 20-27 nucleotide, micro (mi) and short interfering (si)RNAs (Amor *et al.*, 2009). These can silence exogenous or endogenous RNA molecules through the RNA interference (RNAi) pathway or (transcriptionally) through small RNA-directed DNA methylation and recruitment of repressor complexes (Eamens *et al.*, 2008; Turner and Morris, 2011). Much less is known about the regulatory long non-coding (lnc)RNAs, the group to which *COOLAIR* belongs. These are typically defined as being over 200 nucleotides in length but are otherwise difficult to categorise (Ponting *et al.*, 2009). They regulate gene expression by both transcriptional and post-transcriptional mechanisms; can be transcribed from inter- or intragenic regions in sense and antisense directions by different RNA polymerases; may or may not contain a 5' cap or poly(A) tail; and can be *cis*- or *trans*-acting (Ponting *et al.*, 2009).

1.3.2 **FLC** *sense and antisense transcripts are transcribed from the same locus*

FLC sense and antisense transcripts coexist at the same locus. Both are transcribed by RNA Pol II, are spliced and have a 5' cap and poly(A) tail (Swiezewski *et al.*, 2009). *COOLAIR* transcripts downregulate *FLC* sense expression forming part of a self-regulatory circuit. At ambient temperature (pre-vernalization), *FLC* expression levels are set by a chromatin silencing mechanism involving alternative processing of the antisense (in the autonomous pathway). In the cold, *FLC* expression is downregulated by the removal of active chromatin marks, mediated again by the antisense. The exact mechanism(s) of regulation are not known, but are discussed further below.

Antisense transcription is pervasive across plant and animal kingdoms (Yamada *et al.*, 2003; Yelin *et al.*, 2003; Katayama *et al.*, 2005). Overlapping antisense transcripts have

been reported at other protein-coding genes, with some regulating their opposite transcripts in *cis* (Borsani *et al.*, 2005; Henz *et al.*, 2007; Zubko and Meyer, 2007; Earley *et al.*, 2010). Common mechanisms for antisense regulation in *cis* are transcriptional interference (Martens *et al.*, 2004; Hirota *et al.*, 2008) or through RNAi-mediated degradation of their corresponding sense transcripts (Henz *et al.*, 2007). Others are thought to recognise their reverse complementary sequences in RNA or DNA, promoting interaction with protein modifiers to modulate expression (Yap *et al.*, 2010). More recently, antisense transcripts have been thought of as part of self-regulatory circuits, allowing genes to regulate themselves (Pelechano and Steinmetz, 2013).

1.3.3 Antisense regulation through the RNAi pathway

It is possible that *COOLAIR* interacts with sense mRNA transcripts where they overlap (sense exons 1 and 7) and promotes their degradation via an RNAi-mediated pathway. RNAi is the mechanism by which small RNAs bind complementary RNA or DNA targets and induce a silencing effect at a transcriptional or post-transcriptional level. A typical RNAi-based mechanism for miRNA post-transcriptional silencing in plants is as follows. Endogenous precursor long pri-miRNA transcripts are cleaved by a Dicer RNase III enzyme into ~64 nt pre-miRNA stem-loop intermediates and then again into mature ~22 nt miRNAs in the nucleus (Park *et al.*, 2005; Eamens *et al.*, 2008). miRNAs associate with the RNA-Induced Silencing Complex (RISC) in the cytoplasm, where they base pair with complementary regions of target mRNAs to form a double-stranded hybrid which is recognised by an Argonaute protein (normally AGO1) and cleaved (Gregory *et al.*, 2005). mRNA intermediates are consequently degraded and protein expression decreases.

DICER-LIKE (DCL) 1 and 3 enzymes, involved in small RNA generation, are thought to play a role in the autonomous floral promotion pathway (Schmitz *et al.*, 2007). DCL3 is involved in the production of the small 24 nt RNAs that localise to the *COOLAIR* promoter (Swiezewski *et al.*, 2007). Numerous lncRNAs in mammals and plants have been revealed as small RNA precursors, therefore *COOLAIR* lncRNAs are a potential source of these short transcripts (Mourelatos *et al.*, 2002; Allen *et al.*, 2005; Kapranov *et al.*, 2007; Amor *et al.*, 2009; Xin *et al.*, 2011). *FLC* suppressor mutagenesis screens have not brought up any other RNAi pathway components and there is no evidence for *DCL* genes playing a role in vernalization, however (Schmitz *et al.*, 2007). Intact long *COOLAIR* transcripts play a critical regulatory role in the autonomous pathway, therefore RNAi as a primary mechanism for regulation seems unlikely. This does not, however, rule out a role for the small RNAs in targeting FLD activity in the autonomous pathway (Liu *et al.*, 2010). *FLC*

mRNA was found to be highly stable (Csorba *et al.*, 2014), so post-transcriptional interaction with *COOLAIR* is possible. *COOLAIR*, however, physically associates with *FLC* chromatin (Csorba *et al.*, 2014). A role for recruiting protein modifiers to DNA therefore seems more likely than degradation of mRNA.

1.3.4 Antisense regulation through transcriptional interference

The term 'transcriptional interference' is used here to refer to the direct negative impact of one transcriptional activity on another in *cis* (for review, see: Shearwin *et al*., 2005). Distal *COOLAIR* transcripts extend through the sense transcription start site and into the promoter region. It has been hypothesised that sense repression could therefore be linked to promoter interference (Swiezewski *et al.*, 2009). lncRNAs transcribed upstream of the *Drosophila Ultrabithorax* protein-coding gene disrupt its expression in *cis*, potentially via transcriptional interference of promoter regulatory elements (Petruk *et al.*, 2006). The Class II *COOLAIR* transcripts terminate in the *FLC* promoter, but promoter interference seems unlikely because higher levels of Class II are associated with an active state at ambient temperature (Liu *et al.*, 2007). Other possible mechanisms of transcriptional interference are disruption of sense 3' termination, transcript crashing, or competition for RNA Pol II.

Custom tiling arrays challenged the notion of transcriptional crashing at *FLC*; an increase in *COOLAIR* expression does not correlate with a greater abundance of unspliced sense transcripts at the 5' end of *FLC*, i.e. sense transcription is lower but continues to the 3' end (Marquardt *et al.*, 2014). In agreement, single-molecule (sm)RNA FISH revealed that sense and antisense transcripts are transcribed in a mutually exclusive manner at each *FLC* locus within a cell (Rosa *et al.*, 2016). This precludes transcriptional interference via crashing as a mechanism for *COOLAIR* regulation. It further helps to explain the decrease in *FLC* sense transcription during the cold (as *COOLAIR* transcripts are upregulated) (Swiezewski *et al.*, 2009).

1.3.5 Sense and antisense in a self-regulatory circuit

FLC sense and antisense transcript expression profiles are positively correlated, a correlation broken only by cold treatment (Swiezewski *et al.*, 2009). Expression profiling has revealed similar concordant regulation of sense/antisense pairs across the mammalian transcriptome (Katayama *et al.*, 2005). Cotranscriptional coupling during the warm is a consequence of positive feedback mechanisms reinforcing an active or silent *FLC* expression state (Liu *et al.*, 2010; Marquardt *et al.*, 2014). This circuitry was neatly

demonstrated by experiments investigating the effect of a transcription elongation factor component, CDKC;2, on sense and antisense transcription (Wang *et al.*, 2014A). *COOLAIR* transcription is reduced in a *cdkc;2* mutant, this increases *FLC* sense expression which, in turn, increases activity at the locus and upregulates *COOLAIR*. Efficient splicing and polyadenylation of the proximal *COOLAIR* transcript facilitates a repressive state at *FLC* (Marquardt *et al.*, 2014). A cotranscriptional mechanism linking antisense splicing with sense transcript 5' capping has also been suggested (Li *et al.*, 2015). Altered distal splicing was associated with an increased proportion of *FLC* sense transcripts with a 5' cap in the nucleus and higher Pol II occupancy; perhaps then, canonical distal splicing is antagonistic to the *FRI*-mediated 5' capping pathway (Geraldo *et al.*, 2009).

Kinetic coupling provides a link between RNA polymerase elongation rate, chromatin marks, and splicing (Cáceres and Kornblirrt, 2002). Slower elongation at a repressive locus has been associated with higher exon inclusion (de la Mata *et al.*, 2003; Alló *et al.*, 2009). A slower RNA Pol II at an inactive *FLC* may therefore favour proximal small intron splicing and early polyadenylation which, in turn, promote repressive chromatin marks; thus, maintaining a positive feedback loop. Quantitative variation in transcription and elongation rate were indeed found to coincide with the extent of active chromatin marks over the *FLC* gene body (Wu *et al.*, 2016).

FLC sense and antisense transcripts thus seem to exist in a molecular tug of war. This is supported by anticorrelation of sense and antisense transcription at *FLC* loci (Rosa *et al.*, 2016). A similar bidirectional toggle between sense and antisense lncRNAs has been reported at the *FLO11* locus in yeast (Bumgarner *et al.*, 2009; Bumgarner *et al.*, 2012). In this example, the *PWR1* lncRNA acts as an activator and the *ICR1* lncRNA as a repressor of the upstream protein-coding locus. In smRNA FISH experiments, the *PWR1* lncRNA activator was not detected in all *FLO11* on-cells, nor the *ICR1* repressor in all *FLO11* off-cells (Bumgarner *et al.*, 2012). These lncRNAs may therefore be involved in transitioning between (but not maintenance of) the on- or off-states. *COOLAIR* may similarly play a role in the initial setting of *FLC* in an active or repressive state.

Transcription of *COOLAIR* in a higher percentage of cells in the cold (30% versus 3%) explains quantitative increases in *COOLAIR* expression from total RNA analysis (Rosa *et al.*, 2016). Interestingly, intensity and abundance of the fluorescently-tagged *COOLAIR* RNAs also increased during the cold, with dense clouds of antisense transcripts observed around each *FLC* locus (Rosa *et al.*, 2016). An increase in cells expressing *COOLAIR*

reverses cell-size dependent *FLC* mRNA production (Ietswaart *et al.*, 2017). These data support a *cis*-regulatory function for *COOLAIR* in *FLC* regulation during the cold. They also support the idea that *COOLAIR* is retained at and regulates the locus from which it is transcribed, rather than interacting with sense mRNA transcripts or working in *trans*. The positive correlation previously described between *COOLAIR* and *FLC* sense expression profiles could be explained by similar chromatin environments and *trans*-factor concentrations under the same cellular conditions (Rosa *et al.*, 2016).

1.3.6 Antisense regulation through protein recruitment

COOLAIR mediates reduction of H3K36me3 in the nucleation region at *FLC* during cold, and therefore may be responsible for recruiting a H3K36 demethylase or inhibiting a methyltransferase (Csorba *et al.*, 2014). This is supported by opposing H3K36me3 /H3K27me3 profiles in *FLC* alleles with or without the SNP that alters distal splicing (Li *et al.*, 2015). Multiple lncRNAs have been reported to regulate gene expression through interaction with protein complexes or other regulatory elements. For example, the mammalian lncRNAs *MALAT1* and *TUG1* have been reported to alternately interact with CBX4, a component of PRC1, according to its methylation status; *TUG1* binds methylated CBX4 and *MALAT1* nonmethylated (Yang *et al.*, 2011). They can then both act as protein scaffolds, with *MALAT1* interacting with activating proteins and *TUG1* with repressive proteins. *HOTAIR* is another example of a lncRNA thought to act as a protein scaffold, directing PRC2 and chromatin remodelling factors to metastasis suppressor genes in *trans* (Gupta *et al.*, 2010; Tsai *et al.*, 2010). The mammalian *Chaer* lncRNA is thought to do the opposite; it inhibits PRC2-directed repressive chromatin marks by competitively interacting with its catalytic subunit via a 66 nt sequence motif (Wang *et al*., 2016).

RNAs naturally fold into complex secondary and tertiary structures *in vivo* (Ding *et al.*, 2014; Wan *et al.*, 2014; Gosai *et al.*, 2015). lncRNAs can interact with other RNAs, DNA and protein via structural domains. An approximately 400 nt adenine-repeat region in the *Xist* lncRNA folds into complex repeating duplexes which facilitate binding of the epigenetic silencing protein, SPEN (Lu *et al*., 2016). Secondary structure motifs surrounding protein-binding domains in *HOTAIR* are evolutionarily conserved (Somarowthu *et al.*, 2015). *COOLAIR* interacts with *FLC* chromatin, potentially through sequence-specific binding (Csorba *et al*., 2014); it could therefore recruit regulatory proteins through secondary or tertiary structure elements. Some bacterial RNAs act as ‘RNA thermometers’, with temperature-sensitive secondary structures (Kortmann and

Narberhaus, 2012). Given their upregulation in the cold, the structure of *COOLAIR* RNAs could act as thermosensors.

1.3.7 **COOLAIR** *is regulated by a 3' R-loop*

COOLAIR itself is regulated transcriptionally by an R-loop that covers its promoter and first exon (Sun *et al.*, 2013). Nascent *COOLAIR* transcripts form part of the RNA-DNA hybrid which, when stabilised by the plant homeodomain protein AtNDX, inhibits *COOLAIR* transcription (Sun *et al.*, 2013). Stabilisation of the 3' R-loop may promote production of the small 24 and 30 nt RNAs identified by Swiezewski *et al.* (2007). A similar phenomenon has been noted at mammalian gene terminators; 3' R-loops at G-rich terminator elements were found to induce the production of small antisense RNAs and promote RNAi-mediated H3K9me2 formation (Skourti-Stathaki *et al.*, 2014).

Interestingly, another R-loop has been detected at the 5' end, where the distal *COOLAIR* exon overlaps with the sense promoter and first exon (Zhe Wu, Dean lab, unpublished). This supports the hypothesis that *COOLAIR* transcripts themselves drive the creation of R-loops. If an R-loop at the 3' end of *FLC* in the antisense promoter regulates *COOLAIR* transcription, then it is possible the R-loop at the 5' end regulates sense transcription. This provides another link between distal and sense transcription. *FLC* also forms a stable gene loop between its 5' and 3' ends pre-cold, possibly facilitating sense transcription (Crevillén *et al.*, 2013). Disruption of the gene loop coincides with *COOLAIR* upregulation and *FLC* sense transcriptional repression during the cold. A repressive cold-induced gene loop has recently been reported between the *FLC* promoter and 3' end of the first intron (Kim and Sung, 2017). This is thought to be facilitated by two sense lncRNAs, *COLDWRAP* and *COLDAIR*; involvement of *COOLAIR* has not been tested.

1.4 Flowering time control in *Brassica* species

1.4.1 Conservation of flowering time genes

Homologues of *Arabidopsis thaliana* flowering time genes have been found in other members of the Brassicaceae, including *Arabidopsis lyrata* (Kuittinen *et al.*, 2007), in less closely related dicots, such as pea (Foucher *et al.*, 2003), potato (Navarro *et al.*, 2011) and snapdragon (Bradley *et al.*, 1997), and even in monocots, such as wheat (Greenup *et al.*, 2009) and barley (Digel *et al.*, 2015). *FLC* homologues are widespread in flowering plants suggesting a conserved mechanism of floral repression and cold response. Functional conservation of the FLC protein has been reported in *A. lyrata* (Kemi *et al.*, 2013) and radish (Yi *et al.*, 2014), and in less closely related dicots, such as sugar beet (Reeves *et al.*, 2007). It has been conserved but may have a different (or no) role in some species; for example, flowering time of wild rocket is not affected by vernalization, despite containing an *FLC* homologue which is down-regulated by cold and that can delay flowering in an *A. thaliana FRI flc-3* null line (Taylor *et al.*, 2017). *FLC*-like genes were not detected in the legumes *Medicago truncatula*, *Lotus japonicas*, soybean, or pea, despite conservation of other flowering time genes (Hecht *et al*., 2005). Contrary to prior belief, *FLC*-like genes have recently been identified in monocots and were found to be responsive to cold treatment in *Brachypodium* (Ruelens *et al*., 2013).

1.4.2 Evolutionary history of Brassica *species*

The Brassicaceae mustard family comprises over 330 angiosperm genera, including *Arabidopsis* and *Brassica*. *Arabidopsis* and *Brassica* are thought to have diverged from a common ancestor approximately 43 Ma (based on fossil evidence) (Fig. 1.4A; Beilstein *et al.*, 2010). The ancestral *Brassica* under-went a genome-wide triplication event before diverging into the species that exist today (Rana *et al.*, 2004; Parkin *et al.*, 2005; Cheng *et al*., 2013; Xiao *et al.*, 2013). Commercially important *Brassica* crops include *Brassica juncea*, *Brassica napus*, *Brassica oleracea* and *Brassica rapa*; this work focuses on the latter three. *B. rapa* and *B. oleracea* were hybridised to form the allopolyploid, *B. napus*, less than 10 Ka (Fig. 1.4A) (Rana *et al.*, 2004). Similarly, the allopolyploid *Brassica carinata* is thought to be a hybrid of *B. oleracea* and *Brassica nigra*, and *B. juncea* a hybrid of *B. rapa* and *B. nigra* (Nagaharu, 1935). To distinguish the parental origin of genes within these hybrids the *B. rapa* genome is referred to as the A genome, the *B. nigra* genome as the B genome, and the *B. oleracea* genome as the C genome.

Many plant and animal lineages have undergone whole genome duplication events at some point in their evolutionary history (Bomblies and Madlung, 2014). The common ancestor of *Arabidopsis* and *Brassica* probably underwent several successive polyploid events followed by subsequent diploidization (Vision *et al*., 2000; Simillion *et al.*, 2002; Bowers *et al*., 2003). After divergence, the *Arabidopsis* ancestor remained diploid, whereas the triplicated *Brassica* ancestor became polyploid (Parkin *et al.*, 2005). The ancestral *Brassica* would presumably have been hexaploid, with six copies of each chromosome (Lysak *et al.*, 2005). *B. rapa* and *B. oleracea* are therefore considered paleopolyploids, because although they are now diploid they evolved from a hexaploid ancestor. The reversion from polyploid to diploid can occur over time because of gene diversification or loss and chromosome reduction (Lysak *et al.*, 2006). Genomic rearrangements after rediploidization in *Brassica* contributed to their subsequent speciation (Cheng *et al*., 2014).

An allopolyploid is a polyploid in which distinct sets of chromosomes have come from different species as a result of hybridisation, such as in *B. napus*. In contrast, an autopolyploid is where multiple sets of chromosomes have arisen from genome duplication events, such as in the ancestral *Brassica* species. *B. napus* can further be categorised as an allotetraploid or amphidiploid because it contains the diploid chromosome sets of *B. rapa* (AA) and *B. oleracea* (CC), thus becoming AACC. In autopolyploids, multivalent structures can arise during meisosis, when chromosomes pair, synapse and recombine with multiple (rather than one) other chromosomes simultaneously; these are normally selected against over time through reduced crossover rates and lower fertility (Lloyd and Bomblies, 2016). Preference for homologues over homeologues in allopolyploids promotes diploid-like bivalent structures (homologue pairs) during meiosis (Lloyd and Bomblies, 2016). Homologous recombination therefore occurs between the homologous chromosomes (A and A or C and C) in *B. napus* (Udall *et al.*, 2005). An extra level of diversity can be introduced by a lower level of homeologous recombination between the A and C genome chromosomes (Udall *et al.*, 2005). If both these processes of recombination continue, diversity could eventually be reversed as the homeologues homogenise (becoming a segmental allopolyploid) (Udall *et al.*, 2005).

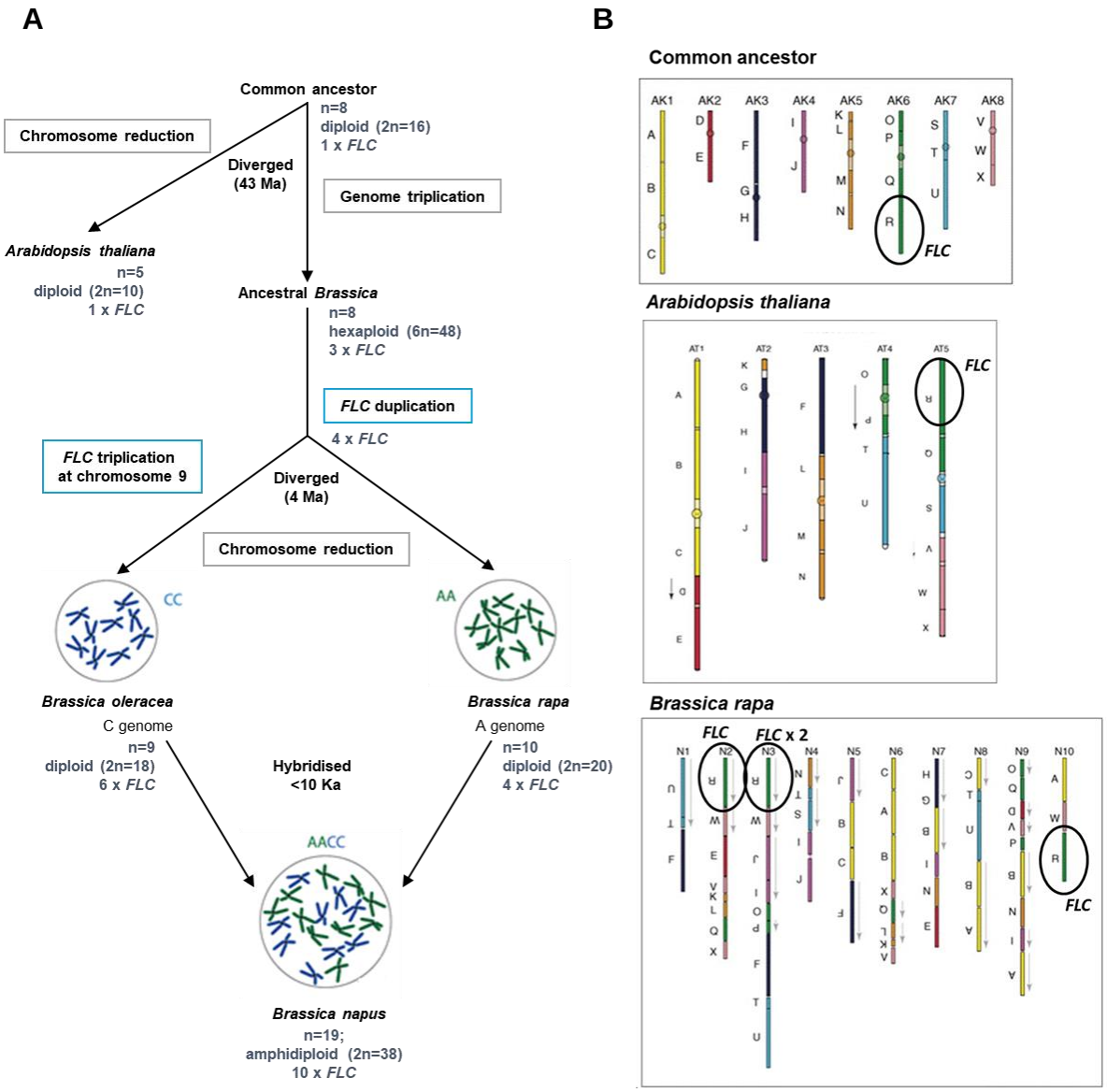


Figure 1.4: Evolution of *Brassica* species. (A) Timeline to represent the major events in the evolution of *Brassica*, with respect to *B. napus*. This has been adapted from Rana *et al.* (2004), with their estimated *Brassica* to *Arabidopsis* divergance date of 20 Ma altered to a more recent estimate of 43 Ma (Beilstein *et al.* 2010). All dates in brackets are approximate. Blue boxes indicate events potentially isolated to a few loci, or perhaps confined to *FLC* alone; grey boxes represent genome-wide events. *Brassica* nuceli schematics are taken from an image by Adenosine at English Wikipedia based on work by Nashville Monkey at English Wikipedia, under a Creative Commons license. (B) Karyotypes of the common ancestor, *A. thaliana* and *B. rapa*, divided into conserved genome blocks, A-X. These schematics were devised by Schranz *et al.* (2006), with full methodology given in their paper. The genome blocks containing *FLC* loci are circled in black.

1.4.3 Genomic context of Arabidopsis *and* Brassica *flowering time genes*

Modern *Arabidopsis* diploids contain duplicated regions of their genome, which are thought to represent up to four duplication events prior to divergence from *Brassica* (Vision *et al.*, 2000; Simillion *et al.*, 2002; Bowers *et al.*, 2003). Since then, both *Arabidopsis* and *Brassica* have undergone vast genome reorganisation and chromosome shuffling (Schranz *et al.*, 2006; Cheng *et al.*, 2013). Conserved chromosome blocks can still be identified within the Brassicaceae and these were used to construct ancestral versus modern-day karyotypes (Fig. 1.4B; Schranz *et al.*, 2006).

A. thaliana contains five chromosomes, which is lower than the average Brassicaceae karyotype and lower than *B. rapa* which has ten (Schranz *et al.*, 2006). The predicted ancestral karyotype is n=8 (where 'n' refers to chromosome number), matched by 37% of its descendants (Lysak *et al.*, 2006). It is thought to have most closely resembled the modern-day *Arabidopsis lyrata* (n=8) and *Capsella rubella* (n=8) genomic structures (Lysak *et al.*, 2006). After speciation, the ancestral Brassicaceae karyotype evolved into the Camelineae karyotype (n=8) and the proto-Calepineae karyotype (n=7) (Murat *et al.*, 2015). *A. thaliana* has therefore under-gone chromosome reduction. Chromosome reduction can occur via a pericentric inversion, where the centromere becomes acrocentric and under-goes reciprocal translocation with the subtelomere of another chromosome (Lysak *et al.*, 2006). This results in a redundant 'mini' chromosome, consisting of centromere and subtelomere, and a much larger fusion chromosome, containing the essential genes (Lysak *et al.*, 2006). The mini chromosomes are meiotically unstable and supposedly lost (Lysak *et al.*, 2006). In this manner, chromosome number can be reduced without gene loss. It is suspected that this is the mechanism by which *Arabidopsis* went from n=8 to n=5, and early *Brassica* species from hexaploid to diploid, without significant gene loss (Schranz *et al.*, 2006).

Schranz *et al.* (2006) described 24 conserved chromosome blocks, A-X, which are thought to have originated from the ancestral karyotype and are shared in all Brassicaceae species. 90% of the *Brassica* genome can be divided into these blocks and matched to syntenic regions in *Arabidopsis thaliana* (Parkin *et al.*, 2005). These blocks have been duplicated and recombined in *B. napus* but retain their structural organisation. Areas of similarity are thus maintained across the Brassicaceae, with chromosome breakage normally occurring around them (Parkin *et al.*, 2005). The predicted *Brassica* genome triplication event suggests that *B. rapa* and *B. oleracea* should contain three copies of every gene found in *Arabidopsis*, and therefore three of each block. Numerous homologous regions have indeed been found but not all are represented equally; this led to the suggestion that isolated duplication events, rather than genome-wide triplication, occurred in the ancestral *Brassica* (Lukens *et al.*, 2004). In opposition to this idea, and in support of ancient genome triplication, 81% of genes are represented by six or more copies in *B. napus*, and 768 genes in *B. rapa* have been mapped to 365 genes in *A. thaliana* (Parkin *et al.*, 2005; Xiao *et al.*, 2013). Genomic fragments from three copies of each block were identified in *B. rapa* by Cheng *et al*. (2013), providing conclusive evidence for whole-genome triplication. The

latter research further characterised the ancestral (pre-triplication) *Brassica* karyotype as n=7, having descended from a younger variant of the proto-Calepineae genome.

FLC in *A. thaliana* was mapped to the top of chromosome 5 (Lee *et al.*, 1994B), and *FRI* to the top of chromosome 4 (Clarke and Dean, 1993). However, in other *Arabidopsis* species, and in the ancestral karyotype, *FLC* is found within block R of chromosome 6 and *FRI* within block W of chromosome 8 (Fig. 1.4B; Schranz *et al.*, 2006). *A. thaliana* chromosome 5 contains parts of the ancestral chromosomes 6, 7 and 8, which have since been reduced and translocated. This explains the current location of *AtFLC*. *FRI*, however, was not found within block W of chromosome 5, as expected, but block O of chromosome 4, which suggests an additional translocation of the *FRI* gene independent of block W. Two *FRI* orthologues have been characterised in *B. oleracea*, *BolC.FRI.a* and *BolC.FRI.b*, and these are syntenic with regions of *A. thaliana* chromosome 5, supporting the idea of translocation from chromosome 5 to chromosome 4 (Irwin *et al.*, 2012). *FRI* has remained in block W in *Brassica* and *FLC* in block R but the duplicated copies can now be found on different chromosomes. Thus, the gene context of *FLC* and *FRI* differs between *A. thaliana* and *Brassica*, with *Brassica FRI* in a gene context more closely resembling that of the ancestral karyotype. Differences in the *AtFRI* versus *Brassica FRI* control of flowering time may result from changes in *cis*/*trans* interactions with other flowering time genes.

1.4.4 FLC *in* Brassica *flowering time control and vernalization response*

Flowering time in *Brassica*, as in *A. thaliana*, is a quantitatively controlled trait (Osborn *et al.*, 1997; Bohoun *et al.*, 1998, Schranz *et al.*, 2002; Long *et al.*, 2007; Lou *et al.*, 2007; Okazaki *et al.*, 2007; Zhao *et al.*, 2010; Zou *et al.*, 2012). Some of the major Quantitative Trait Loci (QTL) from early-late flowering crosses in *B. oleracea* (Bohuon *et al.*, 1998), *B. rapa* and *B. napus* (Osborn *et al.*, 1997) were syntenic with the top of *A. thaliana* chromosome 5, which is where the flowering time control genes *FLC*, *FY*, *CO* and *LFY* are found. Subsequently QTLs have been mapped to individual flowering time genes, including *CO* (Lagercrantz *et al.*, 1996; Okazaki *et al.*, 2007), *FLC* (Kole *et al.*, 2001; Schranz *et al.*, 2002; Okazaki *et al.*, 2007), *FLD* (Long *et al.*, 2007), *VRN1* (Long *et al.*, 2007), *FT* (Long *et al.*, 2007) and *FRI* (Osborn *et al.*, 1997; Wang *et al.*, 2011; Zou *et al.*, 2012).

In the 1990s, vernalization-responsive flowering time loci in *Brassica*, derived from segregating populations of biennials and annuals, were mapped to two major QTLs (Osborn *et al.*, 1997). In *B. rapa* these were designated VFR1 and VFR2, and were

homologous to the VFN1 and VFN2 QTLs identified in *B. napus* (Osborn *et al.*, 1997). VFR2 and VFN2 were found to be collinear with the top of chromosome 5 in *A. thaliana*, where flowering time genes *CO*, *EMF1*, *FY* and *FLC* are located (Osborn *et al.*, 1997). VFR2/VFN2 was homologous with *AtFLC* and corresponded to a *Brassica FLC* gene (Kole *et al.*, 2001). This suggested that allelic variation in *FLC* was a major driver of flowering time variation in *Brassica* as in *Arabidopsis* (Kole *et al.*, 2001). The VFR1/VFN1 QTL had fractionated colinearity with multiple regions in the *A. thaliana* genome, including the top of chromosome 4, where *FRI* and *LD* are located (Osborn *et al.*, 1997).

These experiments suggest similar players are involved in *Brassica* flowering time control. In confirmation of functional conservation of flowering time genes, flowering was delayed when the *B. oleracea BolC.FRI.a* allele was transformed into *A. thaliana* Col-0 (Irwin *et al.*, 2012). Furthermore, late flowering haplotypes of *FRI* were found in *B. napus* winter accessions, and early flowering haplotypes in *B. napus* semi-winter and spring types (Wang *et al.*, 2011). Similarly, higher *FLC* sense transcript levels were detected in biennial late flowering cultivars and lower in annual early flowering cultivars of *B. rapa* (Kole *et al.*, 2001; Kim *et al.*, 2007). In *B. napus*, levels of *FLC* transcript correlated quantitatively with degree of lateness in winter versus spring cultivars (Tadege *et al.*, 2001). This was supported by a later study where overexpressed *B. rapa FLC* constructs from Chinese cabbage delayed flowering (Kim *et al.*, 2007). An *FLC* allele from *Arabidopsis* can function in *Brassica*: Tadege *et al.* (2001) transformed a spring *B. napus* with *AtFLC* and found flowering was delayed quantitatively with protein accumulation. Similarly, the early flowering *A. thaliana* Landsberg *erecta* accession was transformed with five different 35s:*BnFLC* constructs and flowering delayed with each (Tadege *et al.*, 2001). These experiments suggest not just that *FLC* function has been conserved between *Arabidopsis* and *Brassica*, but that the flowering time pathways that converge on *FLC* work insimilar ways.

Vernalization in *Brassica* is likely to work via a similar mechanism. Several groups have recorded a decrease in transgenic and endogenous *FLC* transcript levels during vernalization in annual and biennial *Brassica* cultivars (Kole *et al.*, 2001; Tadege *et al.*, 2001; Lin *et al.*, 2005; Kim *et al.*, 2007; Zou *et al.*, 2012). Degree of *FLC* downregulation (and consequently flowering time) quantitatively reflects length of cold exposure (Kim *et al.*, 2007). This correlates with an increase in *FT* expression levels in the shoot apex,

although not in leaves (Lin *et al.*, 2005). This suggests a similar mechanism whereby *FLC* represses *FT* and other floral integrator genes under non-vernalized conditions.

30% of *Arabidopsis* flowering time genes have more than three paralogues in *Brassica rapa*; paralogues were not found for only 3% (Xiao *et al.*, 2013). This suggests multiple copies of flowering time genes are preferentially retained (compared with other genes) (Schiessl *et al*., 2014). Indeed, circadian clock genes (involved in the photoperiod pathway) were found to be retained at a higher frequency than comparator gene sets in *B. rapa* (Lou *et al.*, 2012). This supports preferential retention of highly networked dose-sensitive genes, such as *FLC*. Gene regulatory networks are difficult to define in *Brassica* because there are often multiple copies of each regulator in a pathway, in addition to multiple copies of downstream targets. In many cases, it is unknown which copy interacts with which regulator, or whether they can interact with all of them.

Xiao *et al.* (2013) reported that 88/201 flowering time genes with expression quantitative trait loci (eQTL)s, genomic loci that contribute to variation in expression of mRNAs, had *A. thaliana* homologues. The *trans* or *cis* nature of regulation of an eQTL can be determined, i.e. we can see whether flowering time genes that are regulated in *cis* in *Arabidopsis* are also regulated in *cis* in *Brassica*. 8/36 of the *A. thaliana* and *Brassica* paralogues analysed had common *trans* eQTLs, including *SOC1*, *FT* and *FLC*, suggesting that these are regulated by the same *trans*-acting factors (Xiao *et al.*, 2013). The *B. rapa FLC* copy *BrFLCA2*, for example, is a candidate *trans*-regulator of *SOC1*, mirroring the relationship between *AtFLC* and *AtSOC1*. 16/36 paralogues had different *trans* QTLs, indicating an alternate pathway of regulation in *Brassica* (Xiao *et al.*, 2013). The *B. rapa FLC* copy, *BrFLCA10*, was *cis*- and *trans*-regulated, *BrFLCA2* and *BrFLCA3a cis*-regulated, and *BrFLCA3b trans*-regulated (Xiao *et al.*, 2013).

1.4.5 The multiple copies of FLC *in* Brassica

Identification and mapping of Brassica FLC *copies*

Whilst *Arabidopsis* has retained a single copy of *FLC*, as found in the ancestral karyotype, *Brassica* species have multiple copies (Table 1.1). The genome triplication event in the ancestral *Brassica*, represented by triplication of block R in Fig. 1.4B, implies that *B. rapa* and *B. oleracea* would each have three copies of *FLC*. In fact, four copies were detected in *B. rapa*: *BrFLCA10*, *BrFLCA2*, *BrFLCA3a* and *BrFLCA3b* (Schranz *et al.*, 2002). Originally three were detected in *B. oleracea*: *BoFLCC9*, *BoFLCC3a* and *BoFLCA3b* (Schranz *et al.*, 2002). A fourth copy, *BoFLCC2*, was later mapped to chromosome 2 with

91% sequence homology to *BrFLCA2*, its A genome homeologue (Lin *et al.*, 2005; Okazaki *et al.*, 2007). *BrFLCA10*, *BrFLCA2*, *BrFLCA3a* and *BrFLCA3b* were mapped to chromosomes 10, 2, 3 and 3, respectively (Schranz *et al.*, 2002). *BoFLCC9*, *BoFLCC2*, *BoFLCC3a* and *BoFLCC3b* were mapped to chromosomes 9, 2, 3 and 3, respectively (Schranz *et al.*, 2002; Okazaki *et al.*, 2007).

	Brassica rapa			*Brassica oleracea*			*Brassica napus*		
Clade	**Name**	**Alternate**	**Chro.**	**Name**	**Alternate**	**Chro.**	**Name**	**Alternate**	**Chro.**
FLC1	*BrFLCA10*	*FLC-A1* *BrFLC1*	10	*BoFLCC9a* *BoFLCC9b* *BoFLCC9c*	*FLC-C1* *BoFLC1*	9	*BnFLCA10* *BnFLCC9a* *BnFLCC9b* *BnFLCC9c*	*BnFLC-A1* *BnFLC-C1.1* *BnFLC-C1.2* *BnFLC-C1.3*	10 19 19 19
FLC2	*BrFLCA2*	*FLC-A2* *BrFLC2*	2	*BoFLCC2*	*FLC-C2* *BoFLC2/4*	2	*BnFLCA2* *BnFLCC2*	*BnFLC-A2* *BnFLC-C2 / BnFLC4*	2 12
FLC3	*BrFLCA3a*	*FLC-A3* *BrFLC3*	3	*BoFLCC3a*	*FLC-C3* *BoFLC3*	3	*BnFLCA3a* *BnFLCC3a*	*BnFLC-A3* *BnFLC-C3*	3 13
FLC5	*BrFLCA3b*	*FLC-A4* *BrFLC5*	3	*BoFLCC3b*	*FLC-C4* *BoFLC5*	3	*BnFLCA3b* *BnFLCC3b*	*BnFLC-A4* *BnFLC-C4*	3 13

Table 1.1: *Brassica FLC* nomenclature. Clade refers to one of four distinct ancestral versions of *FLC* which existed pre-*B. oleracea*/*B. rapa* divergence. Rows represent different homeologues of the same clade (i.e. the descendants of *FLC1*, *2*, *3* or *5*), as they have since diversified post-divergence. These homeologues can be found across (*B. rapa* and *B. oleracea*) and within (*B. napus*) species. Name refers to the name assigned to a copy in this report. Alternate refers to the alternate names that may be used for this copy across the literature (not exhaustive). 'Chro.' refers to the chromosome on which that copy is found.

Phylogenetic analyses have revealed four distinct clades with higher sequence identity and a closer evolutionary relationship: *FLC1*, *FLC2*, *FLC3* and *FLC5* (Okazaki *et al.*, 2007). These clades are represented by homeologues in *B. rapa* and *B. oleracea* (and consequently *B. napus*), and therefore represent four ancestral *FLC* copies, which presumably existed post-genome triplication but pre-divergence (Okazaki *et al.*, 2007). This suggests an additional *FLC* duplication event post-triplication (Okazaki *et al.*, 2007). Neighbour-joining analysis revealed that *FLC3* and *FLC5* are more closely related than the other copies; therefore, it is supposed that this duplication event was of the ancestral *FLC3/5* to generate two copies on chromosome 3 (Zou *et al.*, 2012).

Since four copies of *FLC* were characterised in *B. rapa* and *B. oleracea*, at least eight copies were expected in the *B. napus* hybrid. In one of the earliest studies in this area, Tadege *et al.* (2001) identified five of these using a cDNA library screen, and designated them *FLC1-5*. Whole-genome genetic mapping of a resynthesized *B. napus* later revealed eight potential *FLC* loci, four from the A genome and four from the C genome (Udall *et al.*, 2005). Up to ten copies have now been identified in *B. napus*, with the extra copies pertaining to a duplication (or triplication) of *BoFLCC9* in *B. oleracea* to form *BoFLCC9a*,

BoFLCC9b and potentially also *BoFLCC9c* (Chalhoub *et al.*, 2014; Rachel Wells, personal communication).

Structural characterisation of Brassica FLC *copies*

BrFLC and *BoFLC* copies have similar gene structure to *AtFLC* with seven exons divided by six introns (Schranz *et al.*, 2002; Razi *et al.*, 2008). *BrFLC* coding regions have 81.8-84.6% sequence identity with *AtFLC* (Schranz *et al.*, 2002) and *BoFLC* coding regions 83-88% identity (Razi *et al.*, 2008). The intronic regions are less conserved, with highly variable lengths of introns 1 and 6 (Schranz *et al.*, 2002; Razi *et al.*, 2008). The *BnFLC* copies have the same exon-intron architecture as their *B. rapa* or *B. oleracea* parents, and 95-100% sequence similarity with these in the coding regions (Zou *et al.*, 2012).

The 272 bp promoter region, essential for non-vernalized transcription and vernalization-induced downregulation in *A. thaliana* (Sheldon *et al.*, 2002), is poorly conserved in *B. oleracea*, with only 47-70% sequence identity (Razi *et al.*, 2008). Similarly, conservation was low between copies in *B. napus*, although four conserved *cis* blocks were identified (Zou *et al.*, 2012). The 75 bp region critical for expression pre-vernalization in *Arabidopsis* (Sheldon *et al.*, 2002) was conserved to varying degrees in *B. oleracea* (Razi *et al.*, 2008), and not present at all in *BnFLCC9a* (Zou *et al.*, 2012). There was a more highly conserved 30 bp region within the 75 bp; this contained a putative transcription factor CAAT box binding site (Zou *et al.*, 2012). Some parts of *AtFLC* intron 1 are required for maintenance of *FLC* repression during vernalization (Sheldon *et al.*, 2002) and these were conserved to a higher degree than other intronic regions (Razi *et al.*, 2008). Overall this suggests similar mechanisms of cold-response, but different pre-vernalization (ambient temperature) regulatory mechanisms (Razi *et al.*, 2008).

Functional importance of Brassica FLC *copies*

Multiple functional copies of *FLC* were perhaps retained through evolution to add an extra level of quantitative control to flowering time regulation (Schranz *et al.*, 2002). In accordance with this, five out of five tested *B. napus FLC* copies were functional in *A. thaliana* (Tadege *et al.*, 2001). It is possible that each copy contributes to the control of flowering time in *Brassica* species in a dosage-dependent manner. If this is true, allelic variation within copies should be a strong enhancer of flowering time diversity (Schranz *et al.*, 2002).

The major vernalization-responsive VFN2 and VFR2 QTLs in *B. napus* and *B. rapa* were syntenic to *AtFLC* and to each other (Osborn *et al.*, 1996). These were mapped to chromosome 10, which contains the *FLCA10* copy. A QTL for differences in flowering time of spring cultivars has also been mapped to *FLCA10* (Long *et al.*, 2007). There were no qualitative differences in *FLC1* expression in late versus early flowering lines in a study by Pires *et al.* (2004), but quantitative expression levels may have differed. Indeed, a *cis* eQTL was mapped to *BrFLCA10* in double haploid populations, although it did not colocalise with a QTL in this region (Xiao *et al.*, 2013). Functional conservation is confirmed because overexpression of *BrFLCA10* in an *A. thaliana* Col *flc-null* background delays flowering (Kim *et al.*, 2007). *FLCA10* transcripts also exhibit a conserved vernalization response, with downregulation after 3-5 weeks' cold treatment in *B. napus* and *B. rapa* (Kim *et al.*, 2007; Zou *et al.*, 2012). Vernalization requirement in winter *B. napus* cultivars has been linked to the presence of a Tourist-like MITE (miniature inverted-repeat transposable element) in the upstream region of *BnFLCA10* (Hou *et al.*, 2012). In early flowering spring lines, a G to A mutation in the 5' splice site of *BrFLCA10* intron six triggers three alternative splice forms of the sense transcript and non-functional FLC proteins (Yuan *et al.*, 2009). Genetic variation at *BnFLCA10* was recently proposed as a candidate for the winter-spring split in *B. napus* haplotypes (Schiessl *et al.*, 2017). In summary, *FLCA10* is a strong candidate for flowering time control pre-vernalization in both *B. rapa* and *B. napus*. *FLCC9a/b* has not thus far been considered important in either *B. oleracea* or *B. napus*, although expression levels vary between late and early flowering *B. oleracea* lines (Okazaki *et al.*, 2007). The 75 bp region that is critical for non-vernalized expression of *FLC* in *Arabidopsis* is missing in the *B. napus FLCC9* homologues, suggesting they are regulated differently (Sheldon *et al.*, 2002; Zou *et al.*, 2012).

FLCC2 is an important player in flowering time control in *B. oleracea*. Non-functional *BoFLCC2* proteins were detected in early flowering cabbage and broccoli cultivars (Okazaki *et al.*, 2007), and a conserved vernalization response and inverse relationship with *FT* has been reported (Lin *et al.*, 2005; Ridge *et al.*, 2015). The *BoFLCC2* promoter region and coding sequence were cloned, fused to the *GUS* reporter gene and transformed into *A. thaliana*. Similar expression patterns (as in *B. oleracea*) were reported, with highest expression in the shoot apex (Lin *et al.*, 2005; Okazaki *et al.*, 2007). A strong association between allelic variation in *BoFLC*2 and heading date in cauliflower (Ridge *et al.*, 2015) and purple sprouting broccoli (Irwin *et al.*, 2016) has been reported. Flowering time QTL for late-early flowering crosses were mapped to *BoFLCC2* by Okazaki *et al.* (2007), although not by Razi *et al.* (2008). In the latter study, however, mapping populations were

generated from two parent lines that both contained potentially non-functional *BoFLCC2* alleles.

QTL for flowering time and vernalization response have been mapped to *FLCA2* in *B. rapa* (Schranz *et al.*, 2002; Lou *et al.*, 2007; Zhao *et al.*, 2010). In a study by Xiao *et al.* (2013), *BrFLCA2* QTL colocalised with a strong *cis* eQTL, suggesting that differences in flowering time were a direct consequence of changes in *BrFLCA2* expression level, which in turn was linked to *cis* genetic polymorphism at that locus. *BrFLCA2* transcript levels were consistently found to be higher in *B. rapa* late flowering over early flowering pools (Kim *et al.*, 2007; Zhao *et al.*, 2010), and overexpression results in late flowering in a Col *FRI flc-null* background (Kim *et al.*, 2007). An insertion-deletion mutation at *BrFLCA2*, observed in early flowering oilseeds, results in three alternate non-functional splice forms (Wu *et al.*, 2012A). One of these spliceoforms was previously detected after cold treatment (Zhao *et al.*, 2010). These were linked to flowering time control in early flowering *B. rapa* ecotypes (Wu *et al.*, 2012A) and may be the cause of the strong QTL for flowering time mapping to *BrFLCA2*. *BrFLCA2* levels can be reduced by vernalization, but the response is strongest at the seedling stage (Kim *et al.*, 2007; Zhao *et al.*, 2010). *FLC2* therefore seems to be an important regulator of flowering time and vernalization response in both *B. rapa* and *B. oleracea*; it is interesting to speculate whether they have equal or different roles in *B. napus*, or if one becomes redundant.

Based on the research carried out so far, *FLC3* seems less important for flowering time control although its function as a floral repressor appears to have been conserved. In resynthesized *B. napus* lines, *BnFLCC3a* was lost from early flowering lines and replaced by an extra *BnFLCA3a* copy on chromosome 13 (Pires *et al.*, 2004). Segregation analysis confirmed that the replacement of the C genome copy with the A genome copy was associated with early flowering. This indicates a potential role for *BnFLCC3a* in the delay of flowering. Overexpression of *BrFLCA3a* in a Col *FRI flc-null* background also resulted in late flowering (Kim *et al.*, 2007). Vernalization response has been conserved, with a reduction in *FLC3* transcript levels in *B. napus*, *B. oleracea* and *B. rapa* after cold treatment in vernalization-responsive cultivars (Lin *et al.*, 2005; Kim *et al.*, 2007; Zou *et al.*, 2012). *FLCA3a* and *FLCC3a* were also found to be highly expressed under non-vernalized conditions in *B. napus* leaf tissue (Zou *et al.*, 2012). *BoFLCC3a* has not been mapped to any major QTL, although was found within confidence intervals of a flowering time QTL for late-early flowering mapping populations (Schranz *et al.*, 2002; Okazaki *et al.*, 2007; Razi *et al.*, 2008). A QTL controlling flowering time was mapped to the vicinity

of *BnFLCA3a* by Zou *et al.* (2012), however it did not correlate with expression level differences in the parent lines.

The C genome *FLC5* homeologue (*FLCC3b*) does not appear to be important for flowering time control. *FLCC3b* expression was found to be very low or undetectable in late and early flowering *B. oleracea* cultivars (Okzaki *et al.*, 2007), in resynthesized *B. napus* (Pires *et al.*, 2004) and in natural *B. napus* (Zou *et al.*, 2012). It contains a premature stop codon and so may be non-functional or a pseudogene (Razi *et al.*, 2008; Zou *et al.*, 2012). Despite this, *BoFLC5* was found within the confidence interval of a chromosome 3 QTL, therefore some functional alleles may exist (Razi *et al.*, 2008). In contrast, the A genome *FLC5* homeologue (*BrFLCA3b*) may be more important. Flowering time QTL co-segregated to this region in *B. rapa* (Schranz *et al.*, 2002) and *B. napus* (Zou *et al.*, 2012), although expression was not detected by Pires *et al.* (2004) in resynthesized *B. napus*. In the Zou *et al.* (2012) study, *BnFLCA3b* was the only *FLC* copy that showed differential expression in the parental lines from which the QTL population were obtained, and was therefore considered important. Nucleotide differences in the intronic and promoter regions did not correlate with differences in expression and methylation patterns were similar (Zou *et al.*, 2012). However, differential splicing patterns of the sense transcript between the Tapidor and Ningyou7 parents was observed, with a higher percentage of incorrectly spliced transcripts containing a premature stop codon in Ningyou7 (Zou *et al.*, 2012). *BnFLCA3b* is interesting because it is genomically close to *BnFRI-A3* and the cold-responsive gene *BnCBF-A3* (Zou *et al.*, 2012). It was proposed that these functionally-related genes may have clustered together during translocation events, and are now coexpressed and coregulated.

1.4.6 Commercial importance of flowering time control in Brassica *crops*

Brassica species are commercially important crop plants; they comprise leafy, budding and root vegetables, and oilseeds. *Brassica oleracea* varieties include broccoli, cabbage, kale, cauliflower, kohlrabi and Brussel sprouts. *Brassica rapa* varieties include turnip, Chinese cabbage, bok choy and turnip oilseeds. *Brassica napus* varieties are primarily cultivated for oilseed rape, used to make vegetable oil and biofuel. Winter annual *Brassica* species are useful ‘winter cover’ crops, planted in autumn and harvested in spring to protect and improve soils over winter. Over-wintering also produces leafier varieties with greater yield. Oilseed rape flowers provide the additional benefit of enhancing numbers and diversity of insect pollinators (Carruthers *et al.*, 2017).

The organ that is harvested varies across and within *Brassica* species; for example, the root of the *B. rapa* turnip, the seed of the *B. napus* oilseed rape, and the budding head of the *B. oleracea* broccoli. This means that timing of flowering is critical; the crop should be harvested at the correct developmental stage to give maximum yield. Consistent and predictable flowering times are key as they allow the farmer to plan ahead, rotate crops effectively and take advantage of the seasons. Later or earlier flowering than expected can increase susceptibility to stresses, such as frost or high insect populations, and so cause yield losses. Uniformity of flowering and development is further important for efficient and timely harvest. Changing climate can impact yield; increases in precipitation [over the optimum range] was a critical yield-limiting factor for winter oilseed rape and lower temperatures affected development (Zhang *et al.*, 2017). Climate change modelling predicted that temperature increases of 4.5-5.3 °C would prevent *Arabidopsis halleri*, a perennial relative of *A. thaliana*, from flowering (Satake *et al*., 2013).

Brassica cultivars have been bred with different requirements for, and responses to, cold. *Brassica napus* (which has no wild relatives) consequently contains extensive genetic variation at flowering time genes, including non-synonymous SNPs, insertion-deletions, copy number variation, and presence-absence variation (Schiessl *et al*., 2014; Schiessl *et al*., 2017). Genetic marker techniques have facilitated mapping of important flowering time loci for marker-assisted breeding in *Brassica* oilseeds (Snowdon and Friedt, 2004). This allows for improved selection of parents with suitable allele combinations in crossing programmes and for early identification of lines with desired flowering traits (Jung and Müller, 2009). Spring oilseed rape varieties flower early without vernalization; these are often planted in spring in countries with extreme warm or cold winters. Winter oilseed rape varieties flower late (or not at all) without vernalization; these are often planted in autumn as a cover crop in temperate countries, such as the UK. Having a range of cultivars gives farmers the flexibility to choose when to grow their crop.

1.5 Research summary

1.5.1 Aims

This thesis investigates how sense and antisense transcripts at *FLC* regulate flowering time in *Brassica* species. It aims to improve understanding in three areas: (1) conservation and function of long non-coding (lnc)RNAs, (2) the mechanism of *COOLAIR* regulation of *FLC* sense transcripts, and (3) how regulation of different *FLC* copies affects flowering time in *Brassica* species. The latter would improve selection of desired traits in *Brassica* cultivars. Three commercially important *Brassica* species are utilised: *Brassica napus*, *Brassica oleracea* and *Brassica rapa*.

1.5.2 Objectives

1. *To establish how far* COOLAIR *has been conserved in* Brassica.
   Non-coding RNAs are often less well conserved than their protein-coding counterparts, and have been dismissed by some as transcriptional noise (Ponjavic *et al.*, 2007; Graur, 2013). Research within the Dean lab demonstrates that *COOLAIR* has a role in downregulating sense *FLC* transcripts in *Arabidopsis*. We therefore expect *COOLAIR* to have been conserved in other members of the Brassicaceae. Recent publications have revealed conservation of *COOLAIR* in *Arabis alpina* and *A. lyrata* (Caistaings *et al.*, 2014), and in one of four *FLC* copies in *B. rapa* (Li *et al.*, 2016). It remains to be established whether *COOLAIR* has been conserved at other *FLC* copies in *Brassica*.

2. *To determine how multiple* FLC *homologues regulate flowering time in* Brassica.
   There are four copies of *FLC* in *B. rapa*, five in *B. oleracea*, and ten in *B. napus*. In *A. thaliana*, the FLC protein behaves as a floral repressor; research in *Brassica* suggests that this function has been conserved for at least some of the copies. It remains to be established how far their functions have diversified. It will further be important to establish whether they have a collective influence over flowering time in a dosage-dependent manner and/or whether some homologues are more important than others.

3. *To understand how* COOLAIR *regulates multiple* FLC *homologues in a [palaeo]polyploid*
   There are multiple copies of *FLC* in *Brassica* and therefore potentially multiple copies of *COOLAIR*. It remains to be established whether *COOLAIR* transcription has been conserved in one, multiple or all *FLC* homologues. If the former, can a single dominant

COOLAIR transcript regulate all *FLC* copies in *trans*, or does each *COOLAIR* regulate its own copy in *cis*? Natural flowering time variation has arisen through *COOLAIR* in *Arabidopsis* (Li *et al.*, 2015); how far has *cis* variation at *COOLAIR* influenced flowering time in *Brassica*?

1.5.3 Significance of research

Long non-coding (lnc)RNAs have been variously described as critical regulators or as transcriptional noise (Graur, 2013). Reported low DNA sequence conservation supports the hypothesis that they are not biologically significant (Ponjavic *et al*., 2007). This works investigates these opposing claims through analysis of the evolutionary conservation and functional significance of the lncRNAs found at the *FLC* floral repressor in *Arabidopsis*. Although previous research indicates a role for *COOLAIR* in regulating the sense transcripts at *FLC* in the autonomous floral promotion and vernalization pathways, the exact mechanism of regulation has not been solved. Analysis of *COOLAIR* in other plant species may shed light on this. It will be helpful to understand how *COOLAIR* regulates *FLC* when there are multiple *FLC* copies and (potentially) multiple *COOLAIR*s within a single plant, as in the palaeopolyploids, *B. rapa* and *B. oleracea*, and the polyploid, *B. napus*. Investigating whether natural variation at *COOLAIR* has arisen once (in *Arabidopsis*, as described in Li *et al*., 2015) or multiple times across different plant species will provide insight into its use as an evolutionary tool to finetune flowering time.

B. napus, *B. oleracea* and *B. rapa* are three commercially important crop species. They have been extensively bred to produce varieties with a range of flowering times and consequently a diverse range of edible parts. Breeding new varieties with improved control of flowering will help to reduce unreliable or non-uniform harvests, reducing waste and improving yield (Jung and Müller, 2009). In addition, breeding varieties that can adapt to fluctuations in temperature could be useful for combatting climate change. A long-term goal is to give farmers the flexibility to grow high-yielding crops with predictable flowering times over a range of temperatures.

We are far from fully elucidating flowering time networks in *Brassica*. A critical step in improving flowering time prediction would be to establish how the array of different *FLC* alleles at unique loci work independently and together to determine flowering time. *COOLAIR* provides an extra layer of both complexity and control in the commercial finetuning of flowering time. *COOLAIR* further has the potential to provide a novel tool for selection in marker-assisted breeding or in the engineering of desired phenotypes.

2 Methods

2.1 Bioinformatic analyses

Genomic sequences

Details of *FLC* sequences used in this thesis are given in Table 2.1. Any other *COOLAIR* or *FLC* sequences referred to in the text were sequenced by the author (Emily Hawkes or EH) as part of this project.

Species	Accession or cultivar	Source	Website
A. alpina	FJ543377.1	GenBank	https://www.ncbi.nlm.nih.gov/genbank/
A. lyrata	MN47	Phytozome v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
A. thaliana	Col-0	In-house sequencing	
	Var2-6	In-house sequencing	
B. napus	Cabriolet	In-house sequencing	
	Express	Bayer CropScience	
	PPS02144	Bayer CropScience	
B. oleracea	Alboglabra	Bayer CropScience	
B. rapa	Oleifera	Bayer CropScience	
	R018	In-house sequencing	
C. rubella	Monte Gargano	Phytozome v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
E. salsugineum	Pall.	Phytozome v12.1	https://phytozome.jgi.doe.gov/pz/portal.html

Table 2.1: Sources of *FLC* and *COOLAIR* nucleotide sequence.

Sequence alignment

FLC and *COOLAIR* sequences were aligned by: (1) ClustalW multiple sequence alignment with default parameters using AlignX software (Vector NTI Advance 11.5, Invitrogen) or (2) Muscle multiple sequence alignment using Geneious R7 with 5-10 iterations and default parameters; poorly-aligned sections were realigned using the Geneious alignment tool with default parameters. The selected method is indicated in the text or figure captions where relevant.

Phylogenetic analysis

Phylogenetic trees were created from MUSCLE multiple sequence alignment using the HKY neighbour-joining model in Geneious R7; topologies were tested with 1000 bootstrap replicates.

2.2 Secondary structure probing, prediction and analysis

Note: where experiments in this section were carried out by our collaborators – the Sabonmatsu group at Los Alamos National Laboratory – they are designated below as KS (Karissa Sanbonmatsu) or SH (Scott Hennelly). EH refers to where the experimental procedure was carried out by the thesis author.

Chemical-probing of COOLAIR secondary structure (SH and KS)

SHAPE (Selective 2'-hydroxyl acylation analyzed by primer extension) and CMCT (1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate) probing were carried out by Scott Hennelly as described in Hawkes *et al.* (2016). Briefly, RNA was synthesized using the Standard RNA IVT kit (CELLSCRIPT, USA) for run-off transcription. For SHAPE, folded RNA was probed with 1M7 (1-methyl-7-nitroisatoic anhydride) and for CMCT, with 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate. Parallel RNA samples were treated with DMSO as a blank. Both were reacted for 5 min at 22 °C and precipitated. The modified sites of RNA were analysed by reverse transcription using site-specific 5'-fluorophore-labeled primers and SuperScript® III reverse transcriptase (Life Technologies, USA). The samples, supplemented with the dideoxy terminate sequencing products of Cy3-labeled primer extension, were denatured and loaded on an ABI PRISM® 3100-Avant Genetic Analyzer. In combination with full-length lncRNA analysis, three overlapping fragments covering the *A. thaliana* Col-0 *COOLAIR* distal RNA were probed as in Novikova *et al.* (2013) for shotgun secondary structure determination (3S) analysis. Modular regions were determined by comparison to the full-length RNA; non-modular regions were searched for long-range interactions.

Predictions of COOLAIR secondary structure (EH and KS)

Secondary structure predictions were carried out using the manual strategy developed by Weinberg *et al.* (2007) of initial sequence alignment matching syntenic stretches of higher sequence identity, followed by re-alignment of lower sequence identity regions according to structural domains from a SHAPE-probed structure (in this case, that of the distal *COOLAIR* from *A. thaliana* Col-0). For example, *COOLAIR* helices H8-H9 are antisense to a highly-conserved protein-coding region of *FLC* (containing the MADS box motif) and were therefore used to align homologous sequences across the five species. Next, stretches of sequence flanking H8-H9 were shifted to improve alignment with helices in *A. thaliana*. This was repeated and iterated outwards towards the 5' and 3' ends. KS predicted distal secondary structures for: *A. lyrata*, *C. rubella* (exon 2) and *E. salsugineum*; EH for: *A. alpina*, *C. rubella* (exon 1), all *Brassica* species, and non-probed *Arabidopsis* accessions.

Covariance and conservation analyses (EH and KS)

Covariance and conservation analysis was performed for each predicted structure manually by comparison with the SHAPE-probed *A. thaliana* Col-0 structure (Fig. 3.8), unless stated otherwise (EH). Sequence alignment data was used to support and adjust this analysis. This method can detect shifts in conserved sequence to maintain helices in another part of the secondary structure, and maintenance or extension of helices by insertions. For a consensus secondary structure with covariation analysis (KS), please refer to Fig. 3F in Hawkes *et al.* (2016).

2.3 Plant materials and growth conditions

Plant growth and vernalization conditions – glasshouse

Plants on soil in the glasshouse were watered, maintained and chemically-sprayed (where appropriate) by the Norwich BioScience Institutes (NBI) Horticultural Services team; thanks are given for their dedication and care. *A. thaliana* and *A. lyrata* seeds were sown directly onto soil by the Horticultural Services team and stratified in a vernalization chamber at 5 °C with an 8-hour photoperiod and constant humidity for 3 days. They were then moved into the glasshouse under a long-day 16-hour photoperiod (lit by 600W HPS lamps where supplementary lighting was required) at 18 °C, 70% humidity. After 1-2 weeks' growth, seedlings for vernalization treatment were transferred to short-day conditions (8-hour photoperiod) at 5 °C 70% humidity for 2-12 weeks before return to the warm. Plants were transplanted to larger pots as necessary. Individual plants were staked, bagged after seed set (330 x 180 mm glassine bags), and then harvested. Chaff was separated from seeds by threshing.

Brassica seeds were sown directly onto soil by the Horticultural Services team and grown in the glasshouse at 18 °C 70% humidity under a long-day 16-hour photoperiod (as above). After four weeks' growth, plants for vernalization treatment were transferred to short-day conditions (8-hour photoperiod) at 5 °C constant humidity for 2-12 weeks before return to the warm. Plants were transplanted to larger pots as necessary. Individual plants were staked, bagged before the first open flower (380 x 900 mm microperforated cellophane bags to prevent cross-pollination), and then harvested.

Flowering time was scored as 'days to flower', i.e. total number of days between sowing and the first fully open flower. Where plants were vernalized, the number of days spent in vernalization were subtracted from the total to give 'days to flower' at a comparable growth stage. The mean flowering time and range were calculated for three or more plants

per line or cultivar. Leaf tissue samples were taken from the first fully unfurled new leaf from the apex of the plant and immediately frozen in liquid nitrogen or on dry ice. Where biological replicates were required, these were collected from three different plants of the same line or cultivar.

Plant growth and vernalization conditions – tissue culture

Seeds were sterilised by vapour-phase (as in Clough and Bent, 1998) or by liquid sterilization (see Supplementary Methods) and grown on GM-Glu plant media (1 x Murashige and Skoog salts, 1% glucose, 0.5 mg/L pyridoxine, 0.5 mg/L nicotinic acid, 0.5 mg/L thymidine, 100 mg/L inositol, 0.5 g/L Mes, 0.8% agar, pH 5.7), with selection as required. Plants were grown for 2-3 days in the dark at 4 °C and then transferred to controlled environment rooms (20 °C, 16-hour photoperiod, constant humidity) for at least 7 days. After 7-10 days' growth, plants for vernalization were transferred to short-day conditions (8-hour photoperiod) at 6 °C 70% humidity for 4-10 weeks before return to warm.

Whole plant tissue was sampled from tissue culture plates at NV (without vernalization, 10 days after sowing; in the growth chamber), T0 (after vernalization; in the vernalization chamber), and T10 (10 days after vernalization; in the growth chamber). Plants were transferred from tissue culture to soil for flowering time data and seed collection. Leaf tissue was sampled from plants in soil at T30 (30 days after vernalization; in glasshouse).

Plant materials

Details of plant species and accessions used in this thesis are given in Table 2.2, with further details of the JIC core diversity set (OREGIN) cultivars given in Table 2.3.

Species	Accession or cultivar	Source
A. lyrata	MN47	In-house, Peijin Li
A. thaliana	Col *FRI*	In-house (Dean lab), Julia Qüesta
	Col *FRI flc-2*	In-house (Dean lab), Julia Qüesta
B. napus	Winter lines 1-15	Bayer CropScience (cultivar names confidential)
	JIC winter lines	JIC Core Diversity Set (OREGIN), University of Warwick (see Table 2.3 for names)
	JIC summer lines	JIC Core Diversity Set (OREGIN), University of Warwick (see Table 2.3 for names)
	Express	Bayer CropScience, Chikako Shindo
	Major	Bayer CropScience, Chikako Shindo
	Stellar	Bayer CropScience, Chikako Shindo
	Westar	Bayer CropScience, Chikako Shindo
B. oleracera	DH1012	In-house (Irwin lab), Judith Irwin
B. rapa	Maleksberger	In-house (Irwin lab), Judith Irwin
	Purple Top Milan	In-house (Irwin lab), Judith Irwin
	R018	In-house (Irwin lab), Judith Irwin
C. rubella	Cr22.7	In-house (Ostergaard lab), Nicola Stacey

Table 2.2: Plant materials and sources.

Winter		Spring	
ID	**Cultivar**	**ID**	**Cultivar**
W14	Bolko	S18	Bronowski
W25	Capitol	S31	Ceska
W28	Castille	S48	Drakkar
W29	Catana	S49	Duplo
W39	Coriander	S69	Hanna
W42	Darmor	S83	Karoo
W58	Expert	S108	Monty
W85	Kro	S135	Regent
W87	Les	S155	Stellar
W100	Major	S156	Sur400_024
W102	Matador	S173	Topas
W117	Norin	S175	Tri
W127	Primor	S181	Westar
W150	Slapska Slapy		
W166	Tapidor		
W176	Verona		
Exp.	Express		

Table 2.3: *B. napus* JIC core diversity set cultivars. Seeds were provided by the University of Warwick as part of OREGIN (http://www.herts.ac.uk/oregin/information).

2.4 *COOLAIR* detection and splicing analysis

RNA extraction

Leaf material was homogenised with pestle and mortar in liquid nitrogen or with one to two 3 mm tungsten-carbide beads (Qiagen) by the GenoGrinder® 2010 (SPEX SamplePrep). Total RNA was extracted using the 'Hot phenol method', as in Box *et al.* (2011) with some modifications (see Supplementary Methods for adapted protocol). Quality of the RNA was checked by 0.8% agarose gel electrophoresis, and concentration and quality by measuring the absorbance at 260 nm with a NanoDrop® 1000 spectrophotometer (ThermoScientific). DNA was removed with Ambion® Turbo DNA-free kit (Invitrogen), according to the manufacturer's instructions.

Semi-quantitative followed by nested RT-PCR

~2 µg RNA was reverse transcribed with SuperScript® III (Invitrogen) and gene-specific reverse primers, according to the manufacturer's instructions. cDNA was amplified by touchdown PCR using GoTaq® DNA Polymerase (Promega) according to the manufacturer's instructions, followed by up to two rounds of nested PCRs. Up to three primer pairs were designed per *COOLAIR* target transcript to specifically amplify from a single *FLC* homologue, with up to two sets of nested primers to further increase specificity and the amount of product (Table 2.4 and 2.5). Primers were designed from the first and last proximal or distal exon to amplify all spliceoforms. Typical touchdown nested PCR conditions are given in Table 2.6. Equal amounts (typically 30 ng) of cDNA were used per sample within each experiment and, for *COOLAIR* amplification in *Brassica*, this was

initially amplified and equalised at amplicon level for the *UBC* and/or *GAPDH* reference genes (Table 2.7); thus, the first PCR was semi-quantitative. RT-PCR products were separated by agarose gel electrophoresis and each product cut, gel-purified, cloned and sequenced (as below).

Cloning and sequencing of RT-PCR products

RT-PCR products were gel-purified with Wizard® SV Gel & PCR Clean-Up System (Promega), following the manufacturer's instructions. Purified PCR products were ligated with the pGEM®-T-Easy Vector (Promega), as per the manufacturer's instructions, transformed into DH5α competent cells and selected on LB agar plates with 100 µg/ml ampicillin. 20 µl 1 M IPTG and 40 µl 5% X-gal were added for blue-white screening. Colony PCR with GoTaq® DNA Polymerase (Promega) was performed with M13 forward (5'-GTAAAACGACGGCCAGT-3') and reverse (5'-GGAAACAGCTATGACCATG-3') primers, positive products purified by ExoSap treatment (see Supplementary Methods for details), and sequenced with the BigDye® Terminator v3.1 Sequencing Kit (Applied Biosystems), following the manufacturer's instructions and using a T7 forward primer (5'-TAATACGACTCACTATAGGG-3'). Sequencing reads were carried out by GATC Biotech or Eurofins Genomics.

Spec.	Target	Round	Forward primer (5'-3')	Reverse primer (5'-3')
A. lyrata	*FLC-1 COOLAIR* proximal	PCR 1	GCACTGAGAGAGCCACGTTT	ATGTGGGAGCAGAAGCTGAG
		Nest 1	TGCAAAATAGGCCGTAGGCT	TCCGACAATCTTCCGGTGAC
		Nest 2	TAGGCCGTAGGCTTCTTCAC	GTGACTCTCCCGCTGCTTAT
	FLC-1 COOLAIR distal	PCR 1	ACTGAGAGAGCCACGTTTTCT	ACCTTCTCCAAACGTCGCAA
		Nest 1	TGCAAAATAGGCCGTAGGCT	GTCGCTCTTCTCGTCGTCTC
		Nest 2	GCCGTAGGCTTCTTCACAGT	CTCCTCCGGCGATAAGTACG
B. rapa	*FLCA3a COOLAIR* proximal	PCR 1	TAGACGGTAGGCTTCTTCGC	CGCTATGTATTTGAATGGTTGGT
		Nest 1	CGGTAGGCTTCTTCGCCAT	CGGGAGCCGAAGCTGATAAT
		Nest 2	GAACACACAAGCACACTCACT	CACCTGGACAAATCTCCGACA
	FLCA3a COOLAIR distal	PCR 1	GGCCACGTCTCCGTTACAAA	GCCACTTGAACCGAACCTCT
		Nest 1	TAGACGGTAGGCTTCTTCGC	GGCACAGAGACCACTTGGAG
		Nest 2	AGAACACACAAGCACACTCAC	CGCTTCTCGTTGTCTCCTCC
C. rubella	*FLC COOLAIR* proximal	PCR 1	ATTGTGCCACGCCTCCATT	TGTTTCCAAACCTTAGAACCCCT
		Nest 1	AGCAGAAAGGCTTCTTCACTGT	ACCTGCTGGACAAATCTCCG
		Nest 2	TCACTGTGAAGGAAACACAAAGA	CTCCGACAATCTTCCGGTGA
	FLC COOLAIR distal	PCR 1	CCACGCCTCCATTGCAAAAT	AATTAGGGCACAAGGGGCTC
		Nest 1	ACACACAACATGTCGATGCAA	CATCCGTCGCTCTTCTCGTC
		Nest 2	TGCAAAAACTCTCCCTCGGA	TCTACAGCTTCTCCTCCGGT

Table 2.4: Primers to detect *COOLAIR* from *A. lyrata*, *B. rapa* (*FLCA3a*) and *C. rubella*. 'Spec.' = species. Primers designed by EH.

Species	*FLC*	Target	ID	Round	Forward primer (5'-3')	Reverse primer (5'-3')
B. rapa	A10	Total	1	PCR 1	GACTGCCCAAGTCTCCTTTG	TTCGAGCTTGCTTGTGTATCC
		Proximal	2	PCR 1	AGACTGCCCAAGTCTCCTTT	AGGACAAATCTCCGACATCA
			3	PCR 1	AGACTGCCCAAGTCTCCTTT	GCCGAAGCTGATAATATGGA
			4	PCR 1	GACTGCCCAAGTCTCCTTTG	TGGATGTCTCACCAGGACAA
		Distal	9	PCR 1	GACTGCCCAAGTCTCCTTTG	TCAGCTTTCCGTTCTCTGTG
				Nest 1	AAGCCGTATGCTTCTTCACC	CATCCGTCGCTCTTCTTGT
			10	PCR 1	TTTAACCGAACAGAGGGATA	CTCAGTATCTCCGGCTAGTG
				Nest 1	GAGGGATACACAAGCAAGCTC	AAACCGGACCTCAAGATCAA
			11	PCR1	CAAAGTAAGCCGTATGCTTC	CCGAGGAAGAAAAAGTAGAA
				Nest 1	F primer from 10	R primer from 10
	A2	Proximal	18	PCR 1	AAGATGGCCACGTCTCTCTC	AATAATCTTGCGGGAGCCGA
				Nest 1	TCACCATGAAGCAAACACGAG	CGGGAGCCGAAGCTGATAAA
				Nest 2	N/A	N/A
		Distal	22	PCR 1	AAGATGGCCACGTCTCTCTC	ATCTCCGGCGAGAGTTGAAA
				Nest 1	TCACCATGAAGCAAACACGAG	CCAAACGACGCAATGGTCTC
				Nest 2	ACTCTCTACGATTCTCTCTCAAAG	CGTCGCTCTTCTCGTTGTCT
	A3a	Proximal	12	PCR 1	TTAGACGGTAGGCTTCTTCGC	TGGTGATTTTGAGTTTTTGACCT
				Nest 1	CGGTAGGCTTCTTCGCCAT	CGGGAGCCGAAGCTGATAAT
				Nest 2	GAACACACAAGCACACTCACT	CACCTGGACAAATCTCCGACA
		Distal	16	PCR 1	TAGACGGTAGGCTTCTTCGC	GGCACAGAGACCACTTGGAG
				Nest 1	AGAACACACAAGCACACTCAC	CGCTTCTCGTTGTCTCCTCC
				Nest 2	N/A	N/A
	A3b	Proximal	23	PCR 1	GTCCACGTTTCCGTTGCAAAT	AGTCATAGAAGGCAAAAAGAGAAGT
				Nest 1	ATCCTCCCCGTGAAGCAAAT	TGAGGGAGCCGAAGCTGATA
				Nest 2	CCTCCCCGTGAAGCAAATTC	CTTCCTGTAACTCTCCCGCT
		Distal	25	PCR 1	TTCCGTTGCAAATTGGGCAG	AGGCATCTGTTGGGCTTCTC
				Nest 1	ATCCTCCCCGTGAAGCAAAT	TTGGGCTTCTCGTTGTCTCC
				Nest 2	CCTCCCCGTGAAGCAAATTC	CTCCTCCGGGGATAGGTAGG
B. oleracea	C9a	Proximal	35	PCR 1	CCTCTAGGGAAGGAGTCCCC	TGCGAGCCGAAGCTGATTAT
				Nest 1	GGGAAGGAGTCCCCAAGTCT	CACCTGGACAAATCTCCGACA
				Nest 2	AGCCGTATTCTTCTTCGCCG	ATCTTCCGGTAACGCTCCCA
		Distal	38	PCR 1	GCCTCTAGGGAAGGAGTCCC	GAGACAGAAGCCATGGGGAG
				Nest 1	GAAGGAGTCCCCAAGTCTCC	AAGCTCGTCAGCTTTCCGTT
				Nest 2	AGTCCCCAAGTCTCCTTTGC	TTTCCGTTCTCTGTGACGCA
	C2	Proximal	27	PCR 1	CAGTCTTCACCATGAAGCAAACA	GGAAGTGTCACCTGGACAGA
				Nest 1	TCACCATGAAGCAAACACGA	GTCCGGTAACTCTCCCACTG
				Nest 2	CACCATGAAGCAAACACGAG	TCCCACTGCTTTATTAGCCCC
		Distal	29	PCR 1	CAGTCTTCACCATGAAGCAAACA	GGTTCTCGGAGACAGAAGCC
				Nest 1	CTTCACCATGAAGCAAACACGA	CTCCTCGTTGTCTCAGCCTC
				Nest 2	TCACCATGAAGCAAACACGAG	CGTTTTCGCTTCCGTTCCTC
	C3a	Proximal	32	PCR 1	TAGACGGCTGGCTTCTTCAC	TGGAAGCCAAAATGGCTATGTA
				Nest 1	ACCATGAAGCAAAGAAGAGAGGA	GAAGAATCTTGCGGGAGCCG
				Nest 2	AGAGAGGAAGAACACACAAGCA	CGGGAGCCGAAGCTGATAAT
		Distal	33	PCR 1	TAGACGGCTGGCTTCTTCAC	GGCACAGAGACCTCTTGGAG
				Nest 1	GGCTGGCTTCTTCACCATGA	CGTTGAGAAAGCTCGTCAGC
				Nest 2	AGAGAGGAAGAACACACAAGCA	CTCCGCCGGTGATAAGTATG
	C3b	Proximal	39	PCR 1	CGTTTCTGTTGCAAATTGGGC	CTTGAGGGAGGTGAAGCTGA
				Nest 1	GGGCAGTAAACATCCTCACCA	CTTCCTGTAACTCTCCCGCTG
				Nest 2	GGGCAGTAAACATCCTCACCA	CTGTAACTCTCCCGCTGCTT
		Distal	41	PCR 1	TCTGTTGCAAATTGGGCAG	TACCTTCTGCAAACGACGCA
				Nest 1	GGGCAGTAAACATCCTCACCA	ATCTGTCGGGCTTCTCGTTG
				Nest 2	ACCAGTAGCGAAGGACACAA	TACAGCTTCTCCTCCGGTGA
B. napus	A10	Proximal	43	PCR 1	AGACTGCCCAAGTCTCCTTT	AGAGAGAAGGGAATAGCCGGAA
				Nest 1	CACCGTGAAGCAAAGACAGTT	GAGTAATCTTGTGCGAGCCG
				Nest 2	ACCGTGAAGCAAAGACAGTTCT	CTTCCGGTAACGCTCCCAC
		Distal	45	PCR 1	AGTGGAAACCGGACCTCAAG	AAGCCGTATGCTTCTTCACCG
				Nest 1	ACAGTTCTTTTCTTGTTCGGTGC	TGGAAACCGGACCTCAAGAT
				Nest 2	CCGAACAGAGGGATACACAAG	GCCATGGGGAGGAAGAAACTT
	A3a	Proximal	12	PCR 1	As for *B. rapa* 12 (above)	As for *B. rapa* 12 (above)
				Nest 1		
				Nest 2		
		Distal	17	PCR 1	GGCCACGTCTCCGTTACAAA	GCCACTTGAACCGAACCTCT
				Nest 1	TAGACGGTAGGCTTCTTCGC	GGCACAGAGACCACTTGGAG
				Nest 2	AGAACACACAAGCACACTCAC	CGCTTCTCGTTGTCTCCTCC
	C2	Proximal	27	PCR 1	As for *B. oleracea* 27 (above)	As for *B. oleracea* 27 (above)
				Nest 1		
				Nest 2		

		Distal	29	PCR 1 Nest 1 Nest 2	As for *B. oleracea* 29 (above)	As for *B. oleracea* 29 (above)

Table 2.5: Primers to detect *COOLAIR* from *Brassica FLC* homologues (above and previous page). Primers designed by EH. C9 primers were designed with respect to *FLCC9a*, but specificity is not guaranteed because of high sequence similarity with *FLCC9b* and *FLCC9c*.

Round	Step	Conditions	Cycles
PCR 1	Initial denaturation	95 °C 2 min	1
	Denaturation	95 °C 30 s	3, 3, 3, 23 (**35** total)
	Annealing	68 °C, 65 °C, 62 °C, 59 °C, 56 °C 30 s	
	Elongation	72 °C 1 min	
	Final extension	72 °C 5 min	1
Nest 1	Initial denaturation	95 °C 2 min	1
	Denaturation	95 °C 30 s	5, 5, 20 (**30** total)
	Annealing	65 °C, 60 °C, 55 °C 30 s	
	Elongation	72 °C 1 min	
	Final extension	72 °C 5 min	1
Nest 2	Initial denaturation	95 °C 2 min	1
	Denaturation	95 °C 30 s	5, 5, 25 (**35** total)
	Annealing	65 °C, 60 °C, 55 °C 30 s	
	Elongation	72 °C 1 min	
	Final extension	72 °C 5 min	1

Table 2.6: Touchdown nested PCR conditions for *COOLAIR* detection in *Brassica*.

Species	Target	Forward primer (5'-3')	Reverse primer (5'-3')
Brassica	*UBC*	GATCCACCCACCTCGTGTAG	CTGGAGGGAAGTGAATGGTAAC
	GAPDH	AGAGCCGCTTCCTTCAACATCATT	TGGGCACACGGAAGGACATACC

Table 2.7: Reference primers for semi-quantitative PCR in *Brassica*. Primers provided by Andrew Tallis and Judith Irwin.

3'RACE

500 µg total RNA (per experiment) was extracted from *B. rapa* R018 leaf tissue (as described above). Samples were enriched for polyadenylated RNA using PolyATtract® mRNA Isolation Systems (Promega; following the protocol for 'Small-Scale mRNA Isolation: PolyATtract Systems III and IV') and concentrated by isopropanol precipitation (see Supplementary Methods). RNA ligation-mediated 3'RACE was performed as described previously in Liu *et al.* (2009), with modifications. Briefly, ~1 µg RNA was ligated to the 5' end of the RNA/DNA adaptor (5′-Phos/rUrUrUAACCGCATCCTTCTCTCTACCTACCATTGACCTGTT/-3′ddC; 100 pmol) using T4 RNA ligase (NEB), per the manufacturer's instructions. The ligated RNA was purified with RNeasy® Mini Kit (Qiagen), using the provided RNA clean-up protocol. RNA was reverse transcribed with SuperScript® III (Invitrogen) and the RACE-R1 reverse primer which anneals to the adaptor (Table 2.8). cDNA was used as a template for PCR, followed by two nested PCRs, using GoTaq® DNA Polymerase (Promega) with adaptor-specific reverse primers and *Brassica* gene-specific forward primers (Table 2.8). A 1:8 dilution of the PCR reaction from round 1 was used as the template for round 2, and a 1:8

dilution of the PCR reaction from round 2 as the template for round 3. Typical touchdown PCR conditions are given in Table 2.9. For each round of PCR, a control reaction with just one primer was set up to judge the specificity of the different bands observed. The specific bands from the third round of PCR were gel-purified, ligated into the pGEM®-T-easy vector (Promega), cloned and sequenced as described above, but with the RACE-R3 forward primer as the sequencing primer. Up to 96 positive clones were sequenced per sample to look at a spread of polyadenylation sites.

Target	ID	Round	Forward primer (5'-3')	Reverse primer (5'-3')
BrFLCA10 antisense proximal	RACE-R1	PCR 1	GGGAGCGTTACCGGAAGATT	ACAGGTCAATGGTAGGTAGAGAGA
	RACE-R2	Nest 1	TTATCAGCTTCGGCTCGCAC	GGTAGGTAGAGAGAAGGATGCG
	RACE-R3	Nest 2	TCGGCTCGCACAAGATTACT	AGGTAGAGAGAAGGATGCGG
BrFLCA10 antisense unspliced proximal	RACE-R1	PCR 1	CTCGAGCCGGAGAGAGAGTA	ACAGGTCAATGGTAGGTAGAGAGA
	RACE-R2	Nest 1	ACACCCTTTACAAGGGGATAAT	GGTCAATGGTAGGTAGAGAGAAGG
	RACE-R3	Nest 2	TTCGCCGATTAAAGGTGACT	AGGTAGAGAGAAGGATGCGG
BrFLCA10 antisense distal	RACE-R1	PCR 1	GACGGATGCGTCACAGAGAA	ACAGGTCAATGGTAGGTAGAGAGA
	RACE-R2	Nest 1	CTTCCTCCCCATGGCTTCTG	TGGTAGGTAGAGAGAAGGATGC
	RACE-R3	Nest 2	GAGGTCCGGTTTCCACTAGC	GTAGGTAGAGAGAAGGATGCGG
BrFLCA10 sense	RACE-R1	PCR 1	CTCCGACATCAATCTTCCGGT	ACAGGTCAATGGTAGGTAGAGA
	RACE-R2	Nest 1	TCCCCTTGTAAAGGGTGTACG	GTAGGTAGAGAGAAGGATGCG
	RACE-R3	Nest 2	CGAGAGGCTTCGGGTGTAAA	AGGTAGAGAGAAGGATGCGG
BrFLCA3a sense	RACE-R1	PCR 1	CTCCGACATCAATCTTCCGGT	ACAGGTCAATGGTAGGTAGAGA
	RACE-R2	Nest 1	CCTTGAAGGGTGCACGTTG	ATGGTAGGTAGAGAGAAGGATGC
	RACE-R3	Nest 2	CCCTTGGCTGAGAGATTGTGT	AGGTAGAGAGAAGGATGCGG

Table 2.8: *Brassica FLC* sense and antisense 3'RACE primers. Primers designed by EH.

Round	Step	Conditions	Cycles
PCR 1	Initial denaturation	95 °C 2 min	1
	Denaturation	95 °C 30 s	3, 3, 3, 3, 18 (**30** total)
	Annealing	65 °C, 62 °C, 59 °C, 56 °C, 54 °C 30 s	
	Elongation	72 °C 1 min	
	Final extension	72 °C 5 min	1
Nest 1	Initial denaturation	95 °C 2 min	1
	Denaturation	95 °C 30 s	35
	Annealing	55 °C 30 s	
	Elongation	72 °C 1 min	
	Final extension	72 °C 5 min	1
Nest 2	Initial denaturation	95 °C 2 min	1
	Denaturation	95 °C 30 s	35
	Annealing	55 °C 30 s	
	Elongation	72 °C 1 min	
	Final extension	72 °C 5 min	1

Table 2.9: *Brassica* 3'RACE touchdown and nested PCR conditions.

2.5 Expression analysis by qPCR

FLC *sense qPCR primer design and efficiency testing*

FLC sense qPCR primers were designed over exon-exon boundaries to specifically target spliced mRNA from a single homologue. Specificity was increased by designing primers with homologue-specific SNPs towards their 3' end, preferentially at the penultimate, 3rd or 5th position. Amplicons were 50-200 bp in size. Up to 10 primer pairs were designed for each *FLC* homologue and these tested for specificity by visual analysis of PCR products by

agarose gel electrophoresis, followed by sequencing of single PCR products, followed by analysis of qPCR melt curves. Primer efficiencies were tested by qPCR expression analysis of a four-fold dilution series (with 8+ dilutions) to plot a standard curve (Ct values versus log^{10} of template concentration; for method see Ginzinger, 2002). Primers with efficiencies between 90-105% were selected. Primers were tested in multiple species (*B. rapa*, *B. oleracea*, *B. napus*) in at least one spring and one winter line. The final selected *FLC* primers are given in Table 2.10.

FLC	ID		Primer (5'-3')	Tm	%GC	Position	Size (bp)	%E
A10	q15	F	GTCTCCGCCTCCGGGAA	61	71	E1	120	102
	q16	R	AGCTTTTGACTGACGATCCAA	58	43	E2-3		
C9a	q29	F	GTCTCCGCCTCCAGAAAACT	60	55	E1	115	98
	q30	R	TTGACTGACGATCCAAGGCT	59	50	E2-3		
A2	q41	F	CCTCCGGCAAGCTTTACAAC	59	55	E1	114	102
	q42	R	GAGCTTTTGACTGAAGATCCAGA	58	43	E2-3		
C2	q47	F	GGCTAGCCAGATGGAGAAGAATA	59	48	E6-7	104	91
	q48	R	GTGGGAGAGTTACCGGACAA	59	55	E7		
A3a	q62	F	GCTTGTGGAATCAAATGTCGGT	60	45	E4	116	99
	q63	R	GCTTCAACATTAGTTCTGTCTTCCT	59	40	E4-5		
C3a	q70	F	GACGCAGCGGTCTCGT	60	69	E1	132	94
	q71	R	CAAGGATCCTGACCAGGTTATC	58	50	E1-2		
A3b	q80	F	TCTCTGCGAGGCATCTGTTG	60	55	E1	88	100
	q81	R	GATCTTCTCCAGTCTATCCCCG	59	55	E1-2		
C3b	q88	F	AGGAAAACAGTAGCAGACAAGTTA	58	38	E1	156	97
	q89	R	CCAGTCTATCACCGGAGGAGAA	61	55	E1-2		

Table 2.10: *Brassica FLC* spliced sense qPCR primers. 'E(N)' refers to *FLC* exon in which the primer is positioned; E2-3 means that the primer covers the boundary between exons 2 and 3 (and so on). %E columns refers to % primer efficiency. Primers designed and tested by EH.

COOLAIR *qPCR primer design and efficiency testing*

COOLAIR qPCR primers were designed within exons to specifically target total, proximal, or distal transcripts from a single homologue. Specificity was increased by designing primers with homologue-specific SNPs towards their 3' end, preferentially at the penultimate, 3rd or 5th position. Amplicons were 50-200 bp in size. Five or more primer pairs were designed for each *COOLAIR* transcript per *FLC* homologue and these tested for specificity by visual analysis of PCR products by agarose gel electrophoresis, analysis of qPCR melt curves, and sequencing of qPCR products. Non-diluted *COOLAIR* was detected from total cDNA at 30+ cycles; therefore only 3 dilutions could be used to produce standard curves, afterwhich levels became too low to detect. Primer efficiencies were therefore also tested with plasmid DNA containing *Brassica FLCA3a* and *FLCA10* constructs as the template (see 2.7). Primers with efficiencies between 90-105% were selected. *FLCA3a* and *FLCA10 COOLAIR* primers were tested in *B. rapa*, in the DNA plasmids containing *Brassica FLC* constructs and, for *FLCA10*, in the *A. thaliana FLCA10* transgenic lines. Final selected *COOLAIR* primers are given in Table 2.11; efficient primers for total antisense product were not found for *FLCA10*.

FLC	Target / ID	Name		Primer (5'-3')	Tm	%GC	Pos.	Size	%E
A10	Proximal / A10-P4	q109	F	TCGGCTCGCACAAGATTACT	59	50	E2	125 bp	92
		q110	R	GAGAGAAGGGAATAGCCGGAA	58	52	E2		
	Distal / A10-D2*	q123	F	AGGAAAAGTCGTACTTATCACCG	58	43	E3	74 bp	97
		q124	R	CGTCGCTCTTCTTGTCGTCT	60	55	E3		
	Distal / A10-D4	q129	F	ATGCGTCACAGAGAACGGAA	60	50	E3	161 bp	92
		q130	R	CAAATTAGGGCGCAAAGCAC	59	50	E3		
A3a	Proximal/ A3a-P4	q149	F	TCTTCTTCTCCTTCTGCAAGTT	57	41	E2	155 bp	104
		q150	R	GCCAGGATCGCTATGTATTTG	57	48	E2		
	Distal/ A3a-D3	q163	F	TGGTTTCTGTCTCCAAGTGGT	59	48	E3	78 bp	104
		q164	R	CGCTTTAGTATCTCCGGCCA	60	55	E3		
	Total/ A3a-T1	q171	F	AGAGGGCCACGTCTCCGTTA	63	60	E1	50 bp	104
		q172	R	TCATGGCGAAGAAGCCTACC	60	55	E1		

Table 2.11: *B. rapa* and *B. napus COOLAIR* qPCR primers. 'E(N)' in 'Pos.' (position) column refers to *COOLAIR* exon number, %E column refers to % primer efficiency. * Works only in *B. rapa*, not in *B. napus*. Primers designed and tested by EH.

Expression analysis

Total RNA was extracted from leaf tissue (as above), DNA removed with Ambion® TURBO DNA-free kit (Invitrogen), and ~2 µg RNA used in a 20 µl first stand cDNA synthesis reaction with SuperScript® III (Invitrogen) and gene-specific reverse primers, as per the manufacturer's instructions. After cDNA amplification, the 20 µl reaction was diluted with (typically) 180 µl nuclease-free water, 4 µl of which was used in a 10 µl quantitative RT-PCR with LightCycler® 80 SYBR Green I Master (Roche) reaction mix and a Roche LightCycler® 480II instrument, according to the manufacturer's instructions. 4 µl cDNA per sample was added to 384-well plates with an epMotion® 5070 (Eppendorf) robot, followed by 6 µl of master mix (5 µl 2 x conc. SYBR Green 1 Master reaction mix, 0.25 µl 10 µM forward primer, 0.25 µl 10 µM reverse primer, 0.5 µl nuclease-free water) by hand with an automatic electronic pipette. Standard reaction conditions were 95 °C for 5 min followed by 45 cycles of: 95 °C 15 s, 60 °C 25 s, 72 °C 25 s. Absolute values were quantified by Second Derivative Maximum analysis using the LightCycler® 480 Software Version 1.5 (Roche). Absolute values for target genes were normalised to *A. thaliana UBC* or *B. rapa GAPDH* qPCR reference genes (as appropriate, see Table 2.12) using the $2^{-\Delta\Delta C}{}_{T}$ method (Livak and Schmittgen, 2001). -RT (no reverse transcriptase) controls for each sample were included to check for DNA contamination. For each time point three biological replicates were used, with three technical replicates per biological replicate. Where there was one clearly anomalous sample (compared to the other two), this was removed from the results.

Species	Target		Primer (5'-3')	Source	%E
A. thaliana	*UBC*	F	CTGCGACTCAGGGAATCTTCTAA	Dean lab.	97%
		R	TTGTGCCATTGAATTGAACCC		
B. napus *B. rapa*	*GAPDH*	F	AGAGCCGCTTCCTTCAACATCATT	Xiao *et al.* (2013)	92%
		R	TGGGCACACGGAAGG ACATACC		
	*GAPDH**	F	AGAGCCGCTTCCTTCAACATCATT	As above, but redesigned (nt in red) by Fran Robson to improve efficiency.	97%
		R	TGGGAACACGGAAGGACATTCC		

Table 2.12: Reference primers for qPCR. Note that both GAPDH and GAPDH* primers are used and referred to in this thesis. Efficiencies (%E) tested by EH.

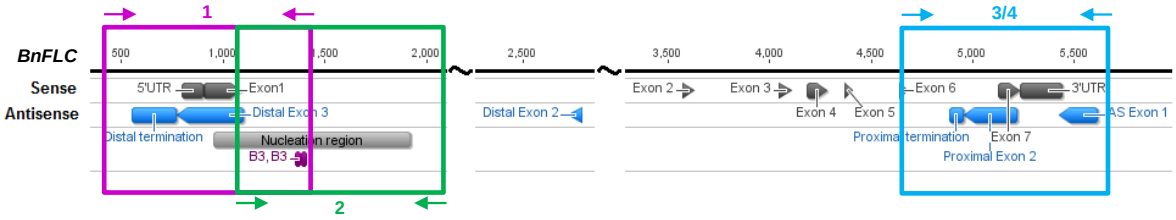
2.6 DNA extraction and sequencing of *COOLAIR*-specific regions

DNA extraction

Leaf tissue was homogenised for DNA extraction by one of three methods: (1) via FastPrep®-24 (Retsch), (2) with pestle and mortar in liquid nitrogen or (3) with disposable grinders. The latter method was used most frequently at JIC, the former was only used at Bayer CropScience (BCS). DNA was extracted by one of two methods: (1) the CTAB protocol, adapted from Murray and Thompson (1980) (see Supplementary Methods), or (2) the Edward's prep, as described in Edwards *et al.* (1991).

Primer design and PCR amplification

Three or more primers were designed per section (1-4) per *FLC* homologue, each to specifically amplify ~1000 bp of target DNA (Fig. 2.1).



ID	Target region	Additional important regions	Coverage	Homologues
1	3' *COOLAIR*	Sense 5'UTR and partial promoter, sense exon 1, partial sense intron 1 and nucleation region, *COOLAIR* distal poly(A) sites	~1000 bp	A10, A3a, C2
2	*Distal COOLAIR* exon	Nucleation region, B3/RY elements, VRE	~1000 bp	A10, A3a, C2
3/4	5' *COOLAIR*	Sense exon 7, sense 3'UTR, sense intron 6, partial *COOLAIR* promoter, *COOLAIR* proximal poly(A) sites	~1000 bp	A10, A3a, C2

Figure 2.1: Target regions 1-4 for sequencing of *COOLAIR* in *B. napus*
Coloured boxes and numbers refer to target regions for sequencing, which cover roughly 1000 bp. Arrows approximate where forward and reverse primers will amplify from. Note that there is an overlap between regions 1 and 2. This schematic was designed with respect to *Brassica rapa FLCA10*; similar exon positions are found for other *FLC* copies and in other *Brassica* species and therefore it works as a basis for all.

Primers were designed from PPS (a spring *B. napus* cultivar, sequence provided by BCS) and checked against Express (a winter *B. napus* cultivar, sequence provided by BCS). Primers were initially tested under a range of annealing temperatures and specificity

checked via 1% agarose gel electrophoresis. Final selected primers and annealing temperatures are given in Table 2.13. 5-30 ng DNA were amplified in the PCR reactions with GoTaq® DNA polymerase (Promega), according to the manufacturer's instructions. Standard reaction conditions were 95 °C 2 min / (95 °C 30 s, Ta 30 s, 72 °C 1 min 30 s) x 35 / 95 °C 5 min, with Ta given in Table 2.13. Three PCR replicates were carried out per sample to distinguish DNA replication errors from polymorphisms (Taq is not proof-reading).

Sequencing

PCR products were purified by ExoSap treatment (see Supplementary Methods for details) and sequenced with the BigDye® Terminator v3.1 Sequencing Kit (Applied Biosystems), following the manufacturer's instructions. PCR products were initially sequenced with both forward and reverse primers from the PCR reaction and, later, with the most efficient of these (or with a purpose-designed sequencing primer) (Table 2.13). Sequencing reads of BCS winter cultivars were carried out by LGC Genomics, and of JIC diversity set cultivars by GATC Biotech.

Target	Sect.	ID		PCR primers (5'-3')	Ta	Sequencing primer (5'-3')
FLCA10	1	EH3	F	AATGCGTATAACTGGTTTTAACTGT	55 °C	EH4
		EH4	R	GCATCACAGCGTGTCAAAGAT		
	2	EH6	F	TTCTTGTCGTCTCCGCCTCC	55 °C	EH6
		EH7	R	TGCCTAAAATATAGTCGCACACTG		
	3	EH12	F	TCCTAGTCTTTTCAGGACCATTTGA	55 °C	EH14
		EH14	R	GGGGTGCTCTCTTGTCCAAA		
FLCA3a	1	EH17	F	ATCTCTCATGAGTCACGGGC	55 °C	EH19
		EH19	R	TCAAAGACAGCTTGAAACTTCACT		
	2	EH23	F	TCTGTGCCTTCCTTTATTAGCCT	63 °C TD	EH24
		EH24	R	GTTCAAGCAGCCGCACATTT		
	3	EH27	F	CTGATGAACATGTTGTCCTTCATA	52 °C	EH28
		EH28	R	TGAAGAGGTCCGTTAAGCGT		
	3*	EH106	F	CATCTGGTCTTTCAGGGATGAT	55 °C	EH107
		EH107	R	AGAGGTCCGTTAAGCGTTGA		
FLCC2	1	EH37	F	TGTCAACAGTTCTGCATTTGTTAAG	63 °C TD	CCAAATCATTTGCATCATCCC (EH111)
		EH34	R	TGCATCACAGGTTATCAAAAAGACC		
	2	EH42	F	CTCAGCCTCTGGCAAGCTAT	52 °C	EH39
		EH39	R	CCCATTCTAATGGCAATCGTGA		
	3	EH100	F	AAGTGTGCTTTATGAGCTCGC	65 °C TD	EH100
		EH101	R	CTCCCAGATCCAGGCTTCAT		
	4	EH51	F	TGGGATGGAGATGCCCTAAA	63 °C TD	EH51
		EH52	R	TGGATGGACTTGCTCTAAGGT		

Table 2.13: *Brassica napus FLC* sequencing primers. 'Ta' refers to the annealing temperature used in PCR amplification. 'TD' refers to where the given Ta was reduced by 3 °C after five cycles two times before 20 more cycles at the final Ta in a touchdown reaction. Sections 1-4 refer to regions of interest within *FLC*. *EH106 and EH107 were used to amplify section 3 from the JIC diversity set (EH27 and EH28 used at Bayer CropScience did not work). Primers designed by EH.

2.7 Synthesis, cloning and transformation of *Brassica FLC* constructs

Sequencing FLCA10 *and* FLCA3a *for construct design*

FLCA10 and *FLCA3a* alleles were sequenced from *B. rapa* Purple Top Milan (PTM) for construct synthesis. PCR primers were designed to cover ~1000 bp overlapping segments of the *FLC* gene, from ~2 kb upstream of the translation start site to ~1.5 kb downstream of the translation stop site (Table 2.14). DNA was extracted from leaf tissue by Edward's prep (Edwards *et al.*, 1991) and amplified by GoTaq® DNA polymerase (Promega), according to the manufacturer's instructions. Standard reaction conditions were 95 °C 2 min / (95 °C 30 s, 63 °C 30 s, 72 °C 1 min 30 s) x 35 / 95 °C 5 min, with Ta reduced by 3 °C for two x 5 cycles in a touchdown reaction. Three PCR replicates were carried out per sample in order to distinguish DNA replication errors from polymorphisms. PCR products were purified by ExoSap treatment (see Supplementary Methods), and sequenced with the BigDye® Terminator v3.1 Sequencing Kit (Applied Biosystems), following the manufacturer's instructions. PCR products were sequenced with both forward and reverse primers from the PCR reaction.

FLCA10				*FLCA3a*			
5'-3'	**ID**		**PCR primers (5'-3')**	**5'-3'**	**ID**		**PCR primers (5'-3')**
1	EH112	F	CTCAAAATCAGTCATAGTCCGCT	1	A3a_1	F	CGATGCAGTAAATGGCAATGAA
	EH113	R	GTCTCCAACAGTGCTTTGCG		A3a_2	R	GGAGGAGACAACGAGAAGCG
2	EH3	F	AATGCGTATAACTGGTTTTAACTGT	2	A3a_3	F	CCGAGTCCACTCGTGATGTC
	EH4	R	GCATCACAGCGTGTCAAAGAT		A3a_4	R	CTCCAAGTGGTCTCTGTGCC
3	EH6	F	TTCTTGTCGTCTCCGCCTCC	3	A3a_3	F	CCGAGTCCACTCGTGATGTC
	EH7	R	TGCCTAAAATATAGTCGCACACTG		A3a_5	R	AAACGGTGCGTTTTGACGAC
4	EH114	F	CCGTGAGGTTATTACTTGATGCT	4	A3a_6	F	ATCTCTCATGAGTCACGGGC
	EH115	R	TTTTGCCCCACACGTTTGTC		A3a_7	R	TCAAAGACAGCTTGAAACTTCACT
5	EH116	F	TAATCGGAAGCGTTGTGTGC	5	A3a_8	F	GGCACAGAGACCACTTGGAG
	EH117	R	TTTTGCCCCACACGTTTGTC		A3a_9	R	GTTCAAGCAGCCGCACATTT
6	EH122	F	TGCCTCTTGGAAACCAACCA	6	A3a_10	F	TCTGTGCCTTCCTTTATTAGCCT
	EH123	R	ACCAGGCTGGAGAGAAGAGA		A3a_11	R	GTTCAAGCAGCCGCACATTT
7	EH124	F	AGGGTTCGTGTGTGTACCTT	7	A3a_12	F	TGCACTGCACTCCGGATAGA
	EH119	R	CTTGTTACGGAGAGGGCGTT		A3a_13	R	CCACGGAGAAGGCAACGAAT
8	EH120	F	TCTCTGCCCTATACATGTTCCA	8	A3a_14	F	ACCGCACATATGCTACACTT
	EH121	R	GCTCATGGTGTGAACCACAG		A3a_15	R	TGTCCACGCTTACACCACC
9	EH125	F	CTGTGGTTCACACCATGAGC	9	A3a_16	F	GTATTGCATTGTTGGTCCACCT
	EH126	R	CTCGAGCCGGAGAGAGAGTA		A3a_17	R	TAGCGATCCTGGCCTCCATA
10	EH12	F	TCCTAGTCTTTTCAGGACCATTTGA	10	A3a_18	F	CTGATGAACATGTTGTCCTTCATA
	EH14	R	GGGGTGCTCTCTTGTCCAAA		A3a_19	R	TGAAGAGGTCCGTTAAGCGT
11	EH127	F	GAGAGAAGGGAATAGCCGGAAA	11	A3a_20	F	AGAATCAGGCTTTGGCTGGC
	EH128	R	AGAGAAAGCGCCAGAGACTG		A3a_21	R	ACCGACTCGGACCCATAGAA
12	EH129	F	TGCGAGCCGAAGCTGATAATA	12	A3a_22	F	TCATGGCGAAGAAGCCTACC
	EH130	R	GCGCCTACGCGGATCAAAAT		A3a_23	R	TGCTTGTAACGTCCCCAAGT
13	EH131	F	TTGCAAAGGAGACTTGGGCA	13	A3a_24	F	ATGGCGAAGAAGCCTACCG
	EH132	R	ACCCGAATACTTCGGTTCGG		A3a_25	R	GTTCGAGAGGATGCAACCCG
				14	A3a_26	F	TGGAACGAGTGGAGGGAAAG
					A3a_27	R	CGGTATCGTACCGGACAAGG

Table 2.14: *Brassica rapa FLCA10* and *FLCA3a* sequencing primers. For amplification from genomic DNA or constructs. Primers designed by EH.

Note that *FLCA10* and *FLCA3a* were originally sequenced in the same manner from *B. rapa* R018, but these were not used for construct synthesis because of an early flowering

mutation in the R018 *FLCA10* allele. *FLCA10* and *FLCA3a* were also sequenced from *B. rapa* Maleksberger at the same time as PTM.

Synthesis and cloning of Brassica rapa FLCA10 *and* FLCA3a *constructs*

B. rapa PTM *FLCA10* and *FLCA3a* constructs with SacI/PstI or SacI/XhoI restriction sites at either end (respectively) were synthesised in the pL0M vector through ENSA (http://www.ensa.ac.uk/). These were digested with SacI/PstI or SacI/XhoI enzymes (NEB, as per their instructions) and gel-purified (Wizard® SV Gel & PCR Clean-Up System (Promega), per manufacturer's instructions). The pSLJ755I6 binary vector was a gift from Prof. Jonathan Jones to the Dean lab (Jones *et al.*, 1992). pSLJ755I6 was linearised in a double digest with the above enzymes and dephosphorylated (with rAPid Alkaline Phosphatase, according to the Roche protocol). Gel-purified constructs were ligated with pSLJ755I6 vectors using T4 DNA Ligase (Promega), as per instructions. These were transformed into DH5α cells and grown under selection on LB agar plates with 10 μg/ml tetracycline. pSLJ755I6 is a low copy number plasmid and so was purified from 100 ml culture with a HiSpeed® Plasmid Midi Kit (Qiagen), as per the provided protocol. Full constructs were then sequenced with forward and reverse sequencing primers (Table 2.14) in a BigDye® Terminator v3.1 (Applied Biosystems) sequencing reaction using the plasmids as template. Sequencing reads were performed by GATC Biotech. Schematics of the constructs and plasmids are given in Fig. 2.2.

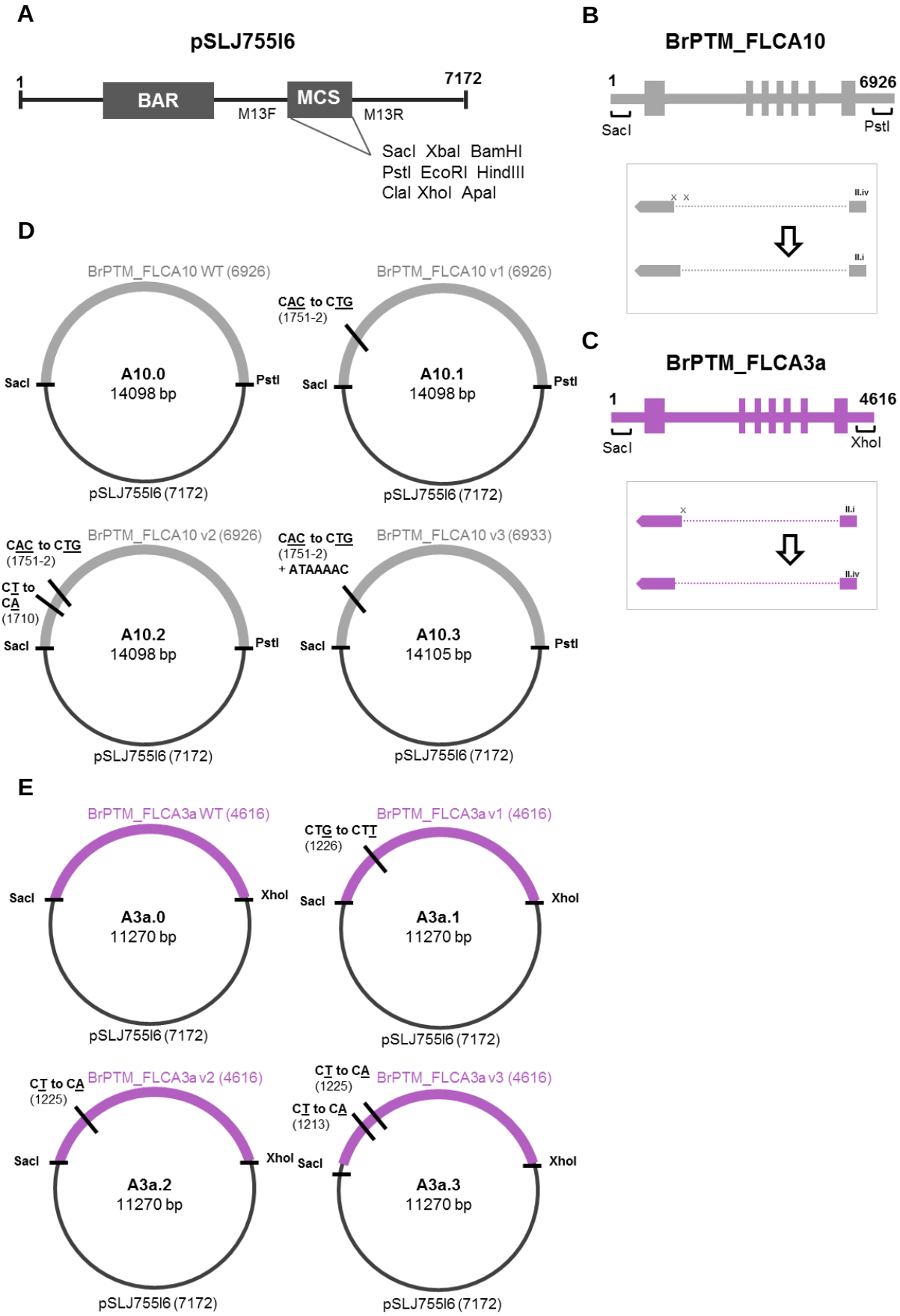


Figure 2.2: Cloning of *B. rapa FLCA10* and *FLCA3a* constructs. (A) Schematic of the pSLJ755I6 vector backbone (Jones *et al.*, 1992). Schematic of *FLCA10* (B) and *FLCA3a* (C) constructs for cloning, with approximate positions of mutated splice sites mapped (x). Final plasmid constructs for *FLCA10* (D) and *FLCA3a* (E) were transformed into *Agrobacterium*. Note that v1-3 refers to the variant *FLC* constructs synthesised, and WT to the wildtype *FLC* construct.

Site-directed mutagenesis and cloning of Brassica oleracea FLCC2 *constructs*

pBlueScript and SLJ1714 plasmids containing the wildtype *FLCC2* construct from the *B. oleracea* PSB E9 allele were previously developed by Clare Lister and Judith Irwin (Irwin *et al.*, 2016). The pSLJ1714 binary vector was a gift from Prof. Jonathan Jones to the Irwin lab (Jones *et al.*, 1992). Primers with desired mutations were designed for site-directed mutagenesis of *BoFLCC2* (Table 2.15). Site-directed mutagenesis was performed with the Q5® Site-Directed Mutagenesis Kit (NEB) according to the manufacturer's instructions, with ~25 ng template plasmid per reaction. Plasmids were transformed into DH5α competent cells, grown under selection on LB agar plates with 100 µg/ml ampicillin, and plasmid DNA prepped with the QIAprep® Spin Miniprep Kit (Qiagen), per the manufacturer's instructions. The full *FLCC2* constructs were then sequenced with sequencing primers (Table 2.16) to confirm that they contained the correct mutation and no other polymorphisms. Plasmid DNA was used as a template in a BigDye Terminator v3.1 (Applied Biosystems) sequencing reaction, with at least two sequencing replicates per sample. Sequencing reads were performed by GATC Biotech.

The six *FLCC2* constructs (wildtype and five variants) in pBlueScript vectors were restricted with XhoI (NEB) and gel-purified (Wizard® SV Gel & PCR Clean-Up System). pSLJ1714 was linearised with XhoI and dephosphorylated with rAPid Alkaline Phosphatase (Roche). *FLCC2* constructs were ligated with pSLJ1714 vectors via their XhoI restriction sites using T4 DNA Ligase (Promega). These were then transformed into DH5α cells and grown under selection on LB agar plates with 10 µg/ml tetracycline. pSLJ1714 is a low copy number plasmid and so was purified from 100 ml culture with a HiSpeed® Plasmid Midi Kit (Qiagen) as per the provided protocol. Orientation was checked in pSLJ1714 vectors by sequencing, and those with forward orientation carried forwards. Schematics of the constructs and plasmids are given in Fig. 2.3.

Target	Variant	ID		Primer (5'-3')	Tm	Mutation
Proximal	P1	SD1	F	AACTAGTGACaAAAAACAAAAATAAGTTAC	57	T to A
		SD2	R	TTGGACTTAAGGGGCTAATAAAG	61	None
	P2	SD3	F	GTGTAAAAACaACGGTTTGATTTGAGTAAAAATATATATTTAAG	63	T to A
		SD4	R	ACGGGGTCTCTCAGCCAA	66	None
Distal	D1	SD7	F	TTCTCTTTACaATTTATAAGAAAATAAAATAAAGTAC	57	T to A
		SD8	R	GAGGAACGGAAGCGAAAAC	61	None
	D2	SD9	F	TTTTTTTATCaGTAAAAGAAATTGCATGTTATTC	58	T to A
		SD10	R	ATCCATGCAGAAAAAGATTG	55	None
	D3	SD11	F	AAAGCTTCTTtCTTTGTTCATTTCTC	55	A to T
		SD12	R	AGCTCATCAGGACATTGC	59	None

Table 2.15: Site-directed mutagenesis primers for *BoFLCC2* constructs. Incorrect nucleotides (the intended mutations) are given in lowercase red. Designed by EH.

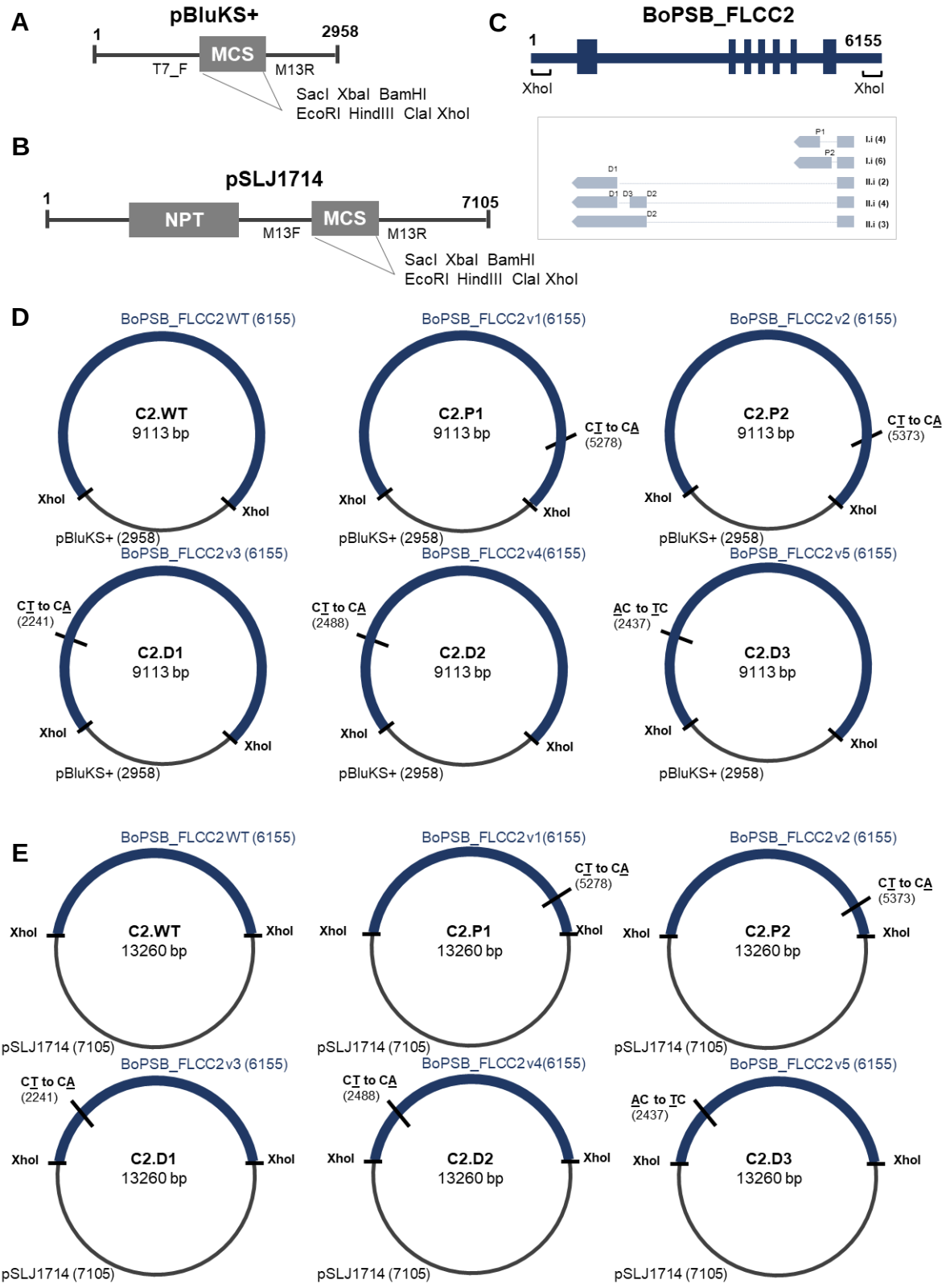


Figure 2.3: Cloning of *B. oleracea FLCC2* constructs. Schematics of pBluKS+ cloning vector backbone (A) and pSLJ1714 binary vector backbone (B). (C) Schematic of *FLCC2* construct for cloning, with blocked splice sites mapped and labelled with construct names (P1, P2, D1, D2, D3). Intermediate (D) and final (E) plasmid constructs; the latter were transformed into *Agrobacterium*. Note that v1-5 refers to the variant *FLC* constructs synthesised, and WT to the wildtype *FLC* construct.

FLCC2			
5'-3'	**ID**		**Sequencing primer (5'-3')**
1	M13F	F	GTAAAACGACGGCCAGT
2	FLC4_F4	F	GAAAAAGTGGGCCAAAATCTCC
3	YWFLC4_F1890	F	AGAGCTGAACCGAACCGAACCT
4	FLC4	R	GTGGAATCAAATTCTGAT
5	FLC4_F5	F	CGCCGGCGATGAGTATG
6	LA11	F	CACATTGTGCAGCCATTAACC
7	AT_FLC4_F4	F	GTAAGCTTGTGGAATCAAATTCTGA
8	FLC4_F2	F	GGCTAGCCAGGTAACAAAG
9	FLC4_F1	F	AGATGCCCTAAAAACCTC
10	FLC4_R1	R	TACAACGCTCACCCTTAT
11	T7	R	TAATACGACTCACTATAGGG

Table 2.16: Sequencing primers for *Brassica oleracea* PSB *FLCC2* constructs. Primers 1 and 11 are designed to amplify from the pBlueScript vector. Primers provided by Clare Lister and Judith Irwin.

Agrobacterium-*mediated floral transformation*

SLJ plasmids containing the desired constructs were transformed into *Agrobacterium tumefaciens* (C58 cells) using the triparental mating method with the helper HB101 plasmid (see Supplementary Methods). Positive transformants were confirmed through colony PCR with M13 forward and *FLC* gene-specific reverse primers. *FLC* constructs were subsequently transformed into Col *FRI flc-2* plants via an *Agrobacterium*-mediated floral spray method, adapted from Clough and Bent (1998). For full details of the modified protocol see Supplementary Methods. 40-80 flowering plants were sprayed two or three times (at five-day intervals) per construct. T0 plants were harvested after seed set.

2.8 Genotyping, copy number analysis and selection of transgenic plants

Sowing and selection of T1 and T2 lines

Hemizygous T1 seeds with *FLCA3a* or *FLCA10* constructs were sown on soil and selected for with BASTA (active ingredient phosphinothricin). Hemizygous T1 seeds with *FLCC2* constructs were liquid sterilised (Supplementary Methods) and sown on GM-Glu plates with 50 μg/ml Kanamycin for selection. Flowering times were recorded at T1 and plants bagged and harvested. Homozygous and heterozygous T2 seeds were vapour-phase sterilised (described in Clough and Bent, 1998) and selected for on GM-Glu plates with 10 μg/ml phosphinothricin (PPT) (*FLCA3a* and *FLCA10*) or 50 μg/ml Kanamycin (*FLCC2*) for selection. Growth conditions given in section 2.3. Flowering times were recorded at T2, leaf tissue sampled for RNA extraction, plants bagged and harvested.

Genotyping and copy number analysis

Leaf tissue was collected from each positive transformant at T1. DNA was extracted by Edward's prep. (Edwards *et al.*, 1991) and genotyped for the *FLC* construct. Genotyping was performed by PCR amplification with construct-specific primers and GoTaq® DNA Polymerase (Promega). Extracted DNA was analysed for copy number by iDNA Genetics.

3 Evolutionary conservation of *COOLAIR* in the Brassicaceae family

3.1 Introduction

Since their discovery, long non-coding RNAs (lncRNAs) have in turn been described as critical regulators of the genome (Mattick, 2011) or as transcriptional noise (Struhl, 2007; Graur *et al.*, 2013). Differential expression of lncRNAs has been reported during plant development (Amor *et al.*, 2009) and in response to abiotic (Amor *et al.*, 2009; Xin *et al.*, 2011; Wu *et al.*, 2012B; Shuai *et al.*, 2014; Wang *et al.*, 2014B; Zhang *et al.*, 2014) and biotic (Xin *et al.*, 2011; Zhu *et al.*, 2014) stress. Examples of lncRNAs with experimentally validated function are limited, with poor sequence conservation calling apparent functionality into question (Wang *et al.*, 2004; Nitsche *et al.*, 2015). This contrasts with the high level of conservation recorded for infrastructural or other well-defined non-coding RNA classes, such as the miRNAs (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee and Ambros, 2001) and snoRNAs (Hüttenhofer *et al.*, 2001). This chapter concerns the functional significance of a group of *Arabidopsis thaliana* antisense lncRNAs (*COOLAIR*) that are differentially regulated during plant development.

lncRNAs have been linked to gene regulation through transcriptional interference (Martens *et al.*, 2004), chromatin remodelling (Hirota *et al.*, 2008), protein complex recruitment (Campalans *et al.*, 2004; Huarte *et al.*, 2010) – potentially as 'scaffolds' (Wierzbicki *et al.*, 2009), target mimicry (Franco-Zorrilla *et al.*, 2007), or as precursors in small RNA-mediated mRNA decay pathways (Allen *et al.*, 2005; Borsani *et al.*, 2005; Kapranov *et al.*, 2007; Zubko and Meyer, 2007). Examples with experimentally-validated functions include *Xist*, which is involved in X chromosome inactivation in mammals, and *HOTAIR*, a regulator of metastasis suppressor genes. The *Xist* lncRNA is expressed from one X chromosome in each cell in XX females; *Xist* spreads along the chromosome to silence transcription and promote a repressive chromatin state (Penny *et al.*, 1996; Galupa and Heard, 2015). *Xist* interacts with proteins, including a Lamin B receptor transmembrane protein, and other lncRNAs, such as *FIRRE* and *ICCE*, to promote nuclear relocalisation and chromosome remodelling of the X chromosome (Chen *et al.*, 2016; da Rocha and Heard, 2017). Although *Xist* has a well-characterized function that is conserved across mammalian species, its nucleotide sequence is poorly conserved with only small stretches of conservation (Nesterova *et al.*, 2001). The mammalian lncRNA *HOTAIR* is transcribed from the *HOXC* locus of homeobox genes (Rinn *et al.*, 2007). It silences target genes at

HOXD and other genomic loci in *trans* by the recruitment of PRC2 and lysine-specific histone demethylase 1 (LSD1) complexes (potentially acting as a protein 'scaffold'), which help to promote a repressive chromatin state (Rinn *et al.*, 2007; Tsai *et al.*, 2011). Although *HOTAIR* is expressed from syntenic loci in humans and mice, its nucleotide sequence and gene architecture are poorly conserved, i.e. it has six exons in humans versus two in mice (Schorderet and Duboule, 2011).

LncRNAs are often less well conserved than protein-coding genes, but more so than introns or intergenic regions (Ponjavic *et al.*, 2007). Low sequence conservation has led to the argument that many (if not most) are transcriptional noise, without biological function (Wang *et al.*, 2004; Struhl, 2007, Graur *et al.*, 2013). This conclusion does not leave scope for ncRNAs with unique patterns of sequence conservation or for positive selection of rapidly evolving ncRNAs (Pang *et al.*, 2006; Kutter *et al.*, 2012; Nitsche *et al.*, 2015). It has been proposed that *HOTAIR* evolved rapidly in mammals, giving rise to different functions in humans and mice (Schorderet and Duboule, 2011). Low sequence constraint would not necessarily impact sense/antisense regulatory mechanisms, nor would it have an impact where variation is matched by covariance at target sequences (Pang *et al.*, 2006). It may be that only short stretches of sequences participate in sequence-dependent functional mechanisms and are therefore constrained (Pang *et al.*, 2006). Furthermore, lncRNAs may function via sequence-independent mechanisms. Diedrichs (2014) describes three further dimensions of lncRNA conservation in addition to primary sequence conservation: conservation of secondary structure, transcriptional conservation and conserved biological function.

Purifying selection can act on secondary structure where implicit in function. Secondary structure describes the first stage of RNA folding – double-stranded helices, hairpins, junctions and single-stranded loops maintained predominantly by non-covalent *A*-form Watson-Crick base pairs (Tinoco and Bustamante, 1999). Less stable tertiary contacts connect secondary structure motifs *in vivo*; these are predominantly non-Watson-Crick base pairs stabilised by divalent ions (Tinoco and Bustamante, 1999). There are a vast number of possible 3D shapes an RNA molecule can fold into; the dominant structure(s) is determined by its stability in the cellular context (thermodynamics) and the kinetics of folding (Tinoco and Bustamante, 1999). RNA folds can be stabilised or altered by protein complex formation or interaction with ligands (Tinoco and Bustamante, 1999). For example, Mg^{2+} ions preferentially bind GU base pairs in secondary structure helix motifs;

these are thought to promote tertiary folding by reducing electrostatic repulsion of the phosphate groups in the RNA backbone (Tinoco and Bustamante, 1999).

There are multiple examples of RNA structure playing an important role *in vivo*; these include the cloverleaf structure of tRNAs (Kim *et al.*, 1974), ribosomal RNA structures (i.e. the 16S rRNA in which functionally important sites are found in non-structured regions, Noller and Woese, 1981) and, more recently, the different structures observed across bacterial riboswitch RNAs (that are thought to be involved in molecular recognition strategies, Roth and Breaker, 2009). A structural motif in the mammalian *Braveheart* lncRNA, which interacts with the CNBP zinc finger protein, is imperative for cardiomyocyte differentiation (Xu *et al.*, 2016). Structural conservation does not have to be linked to sequence conservation; conservation of base pairing properties is more important because reciprocal mutations can maintain the same secondary structure. For example, Smith *et al.* (2013) identified over four million conserved structural elements in mammals, with 88% of these outside known sequence-constrained sites. This included conserved structural elements in *MALAT1*, a lncRNA associated with lung cancer and other human diseases (Gutschner *et al.*, 2013A).

Transcriptional conservation, where production of lncRNAs at syntenic loci have been conserved, can be independent of sequence or structure conservation (Diedrichs, 2014). For example, where the primary function of the lncRNA is to disrupt neighbouring transcripts by transcriptional interference in *cis* (Pelechano and Steinmetz, 2014). lncRNA transcription could also play a role in nuclear architecture; for example, by exposing local chromatin, attracting active chromatin marks and promoting transcription of neighbouring genes (Melé and Rinn, 2016). In these examples, there is likely to be conservation of promoter elements but not of the transcripts themselves. Syntenic lncRNA products without biological function could alternatively be retained if a region of a genome is particularly susceptible to pervasive transcription. Conservation of lncRNA function cannot be assumed (even with all other types of conservation), and must be tested in each species individually. For example, a knockout of the *MALAT1* lncRNA impaired cell migration and formation of metastatic tumour nodules in human cells (Gutschner *et al.*, 2013B) but had no clear phenotype in mice (Eißmann, *et al.*, 2012), despite strong sequence conservation.

COOLAIR lncRNAs are transcribed in the antisense direction at an important *A. thaliana* floral repressor gene, *FLC*, which is quantitatively downregulated during the switch to

flowering. These initiate just downstream of the protein-coding sense transcript poly(A) site and are polyadenylated either at a proximal site to give ~400 nt Class I transcripts or at a distal site within the *FLC* promoter region to give ~750 nt Class II transcripts (Swiezewski *et al.*, 2009; Fig. 3.1). *COOLAIR* are involved in the regulation of the protein-coding *FLC* sense transcripts. Use of alternate *COOLAIR* polyadenylation and splice sites cotranscriptionally regulates *FLC* sense expression and consequently flowering time (Liu *et al.*, 2010; Marquardt *et al.*, 2014). These transcripts act in a feedback mechanism linking *COOLAIR* processing to *FLC* gene body histone demethylation, reduced *FLC* transcription, and earlier flowering (Liu *et al.*, 2010). *COOLAIR* is upregulated during prolonged cold, contributing to a Polycomb-mediated epigenetic switch between opposing chromatin states (Swiezewski *et al.*, 2009; Csorba *et al.*, 2014). The mechanism of regulation by *COOLAIR* is not clear, especially for the distal transcript; it is unknown whether it is *COOLAIR* transcription or the *COOLAIR* transcripts themselves, or both, which are functionally important.

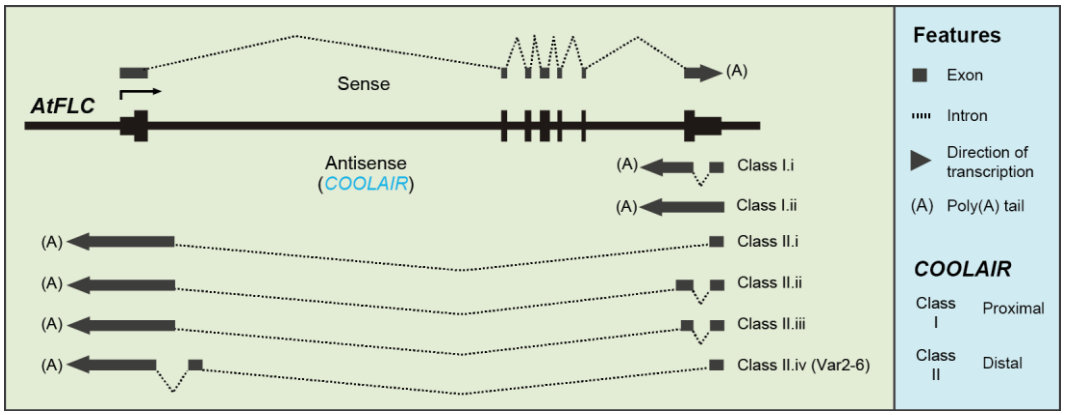


Figure 3.1: Proximal and distal antisense transcripts at *A. thaliana FLC*. *COOLAIR* transcripts have been categorised in *A. thaliana* according to their polyadenylation sites and splicing patterns. They are first divided into Class I (proximal) or Class II (distal) based on where they terminate, and then further divided into Class I.i, Class I.ii (and so on) based on where they are spliced.

Other well-described sense/antisense lncRNA mechanisms include the mammalian *Tsix* lncRNA and the yeast *PWR1* and *ICR1* lncRNAs. *Tsix* is transcribed in the antisense direction to *Xist*; it inhibits *Xist* activation (and therefore X chromosome inactivation) via a *cis*-regulatory mechanism, although whether this is through the act of transcription or the antisense product is (as for *COOLAIR*) unknown (Stavropoulos *et al.*, 2001). Intergenic *PWR1* and *ICR1* are two lncRNAs that are transcribed in opposite directions; they act as a toggle to regulate the *FLO11* locus in yeast, pushing it into an active or silenced state (Bumgarner *et al.*, 2009).

Evolutionary analysis of *COOLAIR* in other *Arabidopsis* accessions and species could provide insight into not only the functional significance of *COOLAIR* and lncRNAs in general, but the mechanism by which they are working. *FLC* homologues are wide-spread in flowering plants and functional conservation of the FLC protein has been reported (Wang *et al.*, 2009B; Guo *et al.*, 2012). Until recently, it was not known whether *COOLAIR* had been conserved in other species. In 2014, Castaings *et al.* reported conservation of *COOLAIR* in *A. alpina* (proximal and distal) and *A. lyrata* (proximal form only), and in 2016, Li *et al.* reported the conservation of *COOLAIR* in one of four *B. rapa FLC* homologues. Here, we present our analyses of *COOLAIR* conservation with respect to three of the four dimensions of lncRNA conservation: primary nucleotide sequence, secondary structure and transcription. We focus on five members of the Brassicaceae family to which *Arabidopsis thaliana* belongs, representing ca. 13-43 million years divergence from *A. thaliana* (Koch and Kiefer, 2005; Beilstein *et al.*, 2010). With regards to nucleotide sequence conservation, the *COOLAIR* promoter region was reported to be evolutionarily conserved but not regions corresponding to *FLC* introns or 5' and 3' UTRs (Castaings *et al.*, 2014; Li *et al.*, 2016). We similarly observed low nucleotide conservation in *COOLAIR*-specific regions, which made it difficult to predict regions of syntenic transcription. We therefore predicted secondary structures as an alternative approach to predict *COOLAIR* exons, which were then validated *in vivo*. This, in conjunction with other recently published work, suggests *COOLAIR* has been evolutionarily conserved and supports its functional significance.

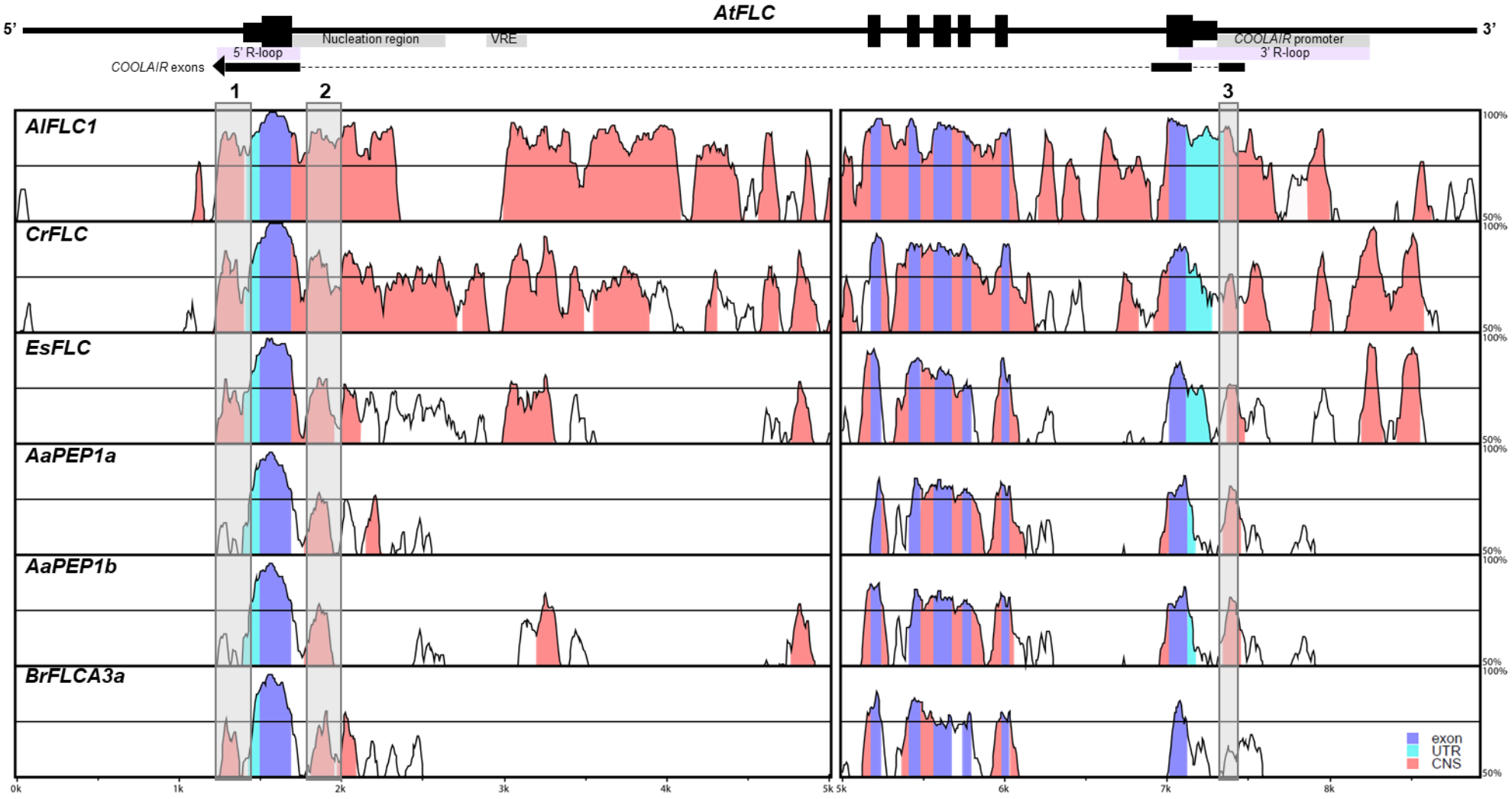
3.2 Results

3.2.1 There is a conserved CG-rich sequence block within COOLAIR exon 1

FLC orthologues in *A. lyrata, C. rubella, E. salsugineum* and *B. rapa* exhibit similar gene structure to *FLC* in *A. thaliana*, with seven exons and six introns. The largest intron (intron 1) divides exons 1 and 2. *PEP1* in *A. alpina* contains a tandem duplication of part of the promoter, first exon and intron 1, but retains similar gene structure from exon 2 to 7 (Albani *et al*., 2012). *PEP1* expresses transcripts from each of the transcriptional start sites (*PEP1a* and *PEP1b*), both splicing to the second exon and producing an apparently functional protein. *A. lyrata* underwent a tandem duplication at the *FLC* locus giving rise to two homologues: *FLC1* (used in the following analyses) and *FLC2* (Kemi *et al.*, 2013). *B. rapa* underwent ancient genome triplication and a subsequent duplication event, giving rise to four homologues at unique loci: *FLCA10, FLCA2, FLCA3a* (used in the following analyses), and *FLCA3b*.

FLC homologues from each of the five species, from 1.5 kb upstream of the translational start site to 1.8 kb downstream of the translational stop site, were aligned with *A. thaliana FLC* to compare primary DNA sequence (Fig. 3.2). Greater sequence homology (coloured blocks of 75% or over) can be observed between *A. thaliana* and *A. lyrata*, and between *A. thaliana* and *C. rubella*, than between *A. thaliana* and the other species. This supports current scientific thinking that *A. lyrata* and *C. rubella* have more recently diverged from *A. thaliana* than the other species in this study (Belstein *et al.*, 2010). Strong blocks of homology are present in all six species at the seven protein-coding exons (marked in dark blue) and, to a lesser extent, at the 5'UTR (light blue). Evolutionary conservation of protein-coding regions supports the biological significance of these *FLC* orthologues. The greatest homology is observed for *FLC* exon 1, which contains the functional MADS box motif (Michaels and Amasino, 1999).

Figure 3.2: Sequence conservation of *A. thaliana FLC* with five Brassicaceae orthologues (next page). MLAGAN progressive pair-wise alignments were performed and visualized with the mVISTA online tool, with a window size of 100 bp and cut-off criteria of 70% sequence identity (Brudno *et al*. 2003; Frazer *et al*., 2004). Coloured areas indicate stretches of homology greater than 75% identity at the nucleotide level; sequence identity (%) is shown on the Y-axis. *FLC* sequences are from 1.5 kb upstream of the ATG start to 1.8 kb downstream of the translational stop. *AaPEP1a* starts 1.5kb upstream of the first *PEP1* ATG start and *AaPEP1b* 1.5 kb upstream of the second ATG start. Numbered grey boxes highlight areas of interest referred to in the text. CNS = conserved non-coding sequence; VRE = Vernalization Response Element.


AtFLC
5'
3'
Nucleation region
VRE
COOLAIR promoter
5' R-loop
3' R-loop
COOLAIR exons
1
2
3
AlFLC1
CrFLC
EsFLC
AaPEP1a
AaPEP1b
BrFLCA3a
100%
50%
exon
UTR
CNS
0k
1k
2k
3k
4k
5k
5k
6k
7k
8k

Aside from a short (~200 bp) block of homology immediately upstream of the 5'UTR in all species except for *A. alpina* (grey box 1, Fig. 3.2), the *FLC* promoter region is devoid of sequence conservation. This suggests that conserved promoter elements are either present immediately upstream of the *FLC* translational start site or further upstream of the 1.5 kb region included in this analysis. Indeed, Castaings *et al.* (2014) reported a conserved sequence block ~2.4-2.8 kb upstream of the *FLC* transcriptional start site. The 3'UTR is less well conserved in *A. alpina* and *B. rapa* than the 5'UTR. There are up to five conserved blocks downstream of the 3'UTR in *A. lyrata*, *C. rubella* and *E. salsugineum*; all but one of these is lost in *A. alpina* and none are observed in *B. rapa*. A ~200 bp block of sequence homology is present at the 5' end of sense intron 1 (grey box 2, Fig. 3.2); this is within the nucleation region and includes B3 protein-binding domains (RY motifs), important for downregulation of *FLC* in the cold (Sheldon *et al.*, 2002; De Lucia *et al.*, 2008; Qüesta *et al.*, 2016). The Vernalization Response Element (VRE) is further downstream of this. The 5' end of the VRE pertains to the *COLDAIR* promoter (Heo and Sung, 2011); this was previously reported to be the least well-conserved part (Castaings *et al.*, 2014).

The ~100 bp conserved block immediately downstream of the 3'UTR pertains to the *COOLAIR* proximal and distal exon 1 (grey box 3, Fig. 3.2). This is interesting because it is within the *COOLAIR* promoter, and is where we see AtNDX binding to stabilise the R-loop and regulate *COOLAIR* expression (Sun *et al.*, 2013). We do not, however, observe this conserved block in the more distantly related *B. rapa FLCA3a*, suggesting evolutionary divergence and lack of sequence conservation in this species. The *COOLAIR* promoter region is thought to extend another 770 bp upstream of this antisense exon (i.e. downstream of the sense 3'UTR). We observe decreasing conservation of the rest of this promoter region as evolutionary distance increases. Exon 2 of the proximal overlaps and extends upstream of *FLC* sense exon 7; the non-overlapping region does not show strong sequence homology across the six species. Sequence conservation of [non-sense-overlapping] distal exons is also low. The afore-mentioned ~200 bp block of sequence homology immediately upstream of the 5'UTR could represent conservation of the distal second exon (grey box 1, Fig. 3.2). Alternatively, this block may contain conserved sense transcript promoter elements. Overall, there appears to be lower sequence conservation of *COOLAIR* exons than of protein-coding exons. This suggests weak purifying selection of total *COOLAIR* transcripts, with the highest selection pressure at *COOLAIR* exon 1. As this exon is within the *COOLAIR* promoter region it is possible that the presence or

absence of *COOLAIR* transcripts is biologically significant, but their nucleotide sequences less so.

Two R-loops are present at the 5' and 3' ends of the *FLC* locus. The 5' R-loop forms from nucleotide sequences pertaining to *FLC* sense exon 1 and the distal *COOLAIR* terminal exon (Zhe Wu, Dean lab, unpublished). The 3' R-loop forms from sequences downstream of the 3'UTR, from the *COOLAIR* promoter and exon 1 region (Sun *et al.*, 2013). R-loops are a three-stranded nucleic acid structure formed by an RNA-DNA hybrid plus a displaced ssDNA strand. They can form *in vivo* during transcription when a nascent RNA transcript invades the dsDNA. The 3' R-loop is ~300-700 bp long, initiating ~200 bp upstream of the *COOLAIR* transcription start sites and terminating 100-500 bp downstream of this (up to the proximal polyadenylation site) (Sun *et al.*, 2013). Low nucleosome density and high GC base pairing provided initial evidence that there was an R-loop at the *COOLAIR* promoter as these features promote R-loop formation (Sun *et al.*, 2013).

GC content was analysed for the six species in this study to determine whether the high %GC region at the *COOLAIR* promoter (and potentially the 3' R-loop) has been conserved (Fig. 3.3). This study reports lower %GC in the 3' region (~50%) in *A. thaliana* than observed by Sun *et al.* (2013); this is still higher than the average ~36% GC content across the *A. thaliana* genome (Meister, 2005). There is a conserved peak across all six species around the translational stop site and the start of proximal exon 2. There is also a conserved ~100 bp region of above average (> 40%) GC content downstream of this, within the *COOLAIR* promoter and coinciding with *COOLAIR* exon 1. This corresponds with the conserved ~100 bp block identified at the *COOLAIR* promoter in five out of six species in Fig. 3.2 (box 3), and may represent conservation of the 3' R-loop. Interestingly, this GC peak is present in *B. rapa*, despite lack of the conserved sequence block. There is also a conserved peak – 50-60% GC content – corresponding to sense exon 1 and the distal terminal exon at the 5' end of *FLC*, although GC content is lower in *B. rapa* and *A. alpina*. This may represent conservation of the 5' R-loop. Other GC peaks are less well conserved across species. Two adjacent peaks of GC > 50% are observed in *A. thaliana*, *C. rubella* and *E. salsugineum* at the 3' end of *FLC*, but not in the other species.

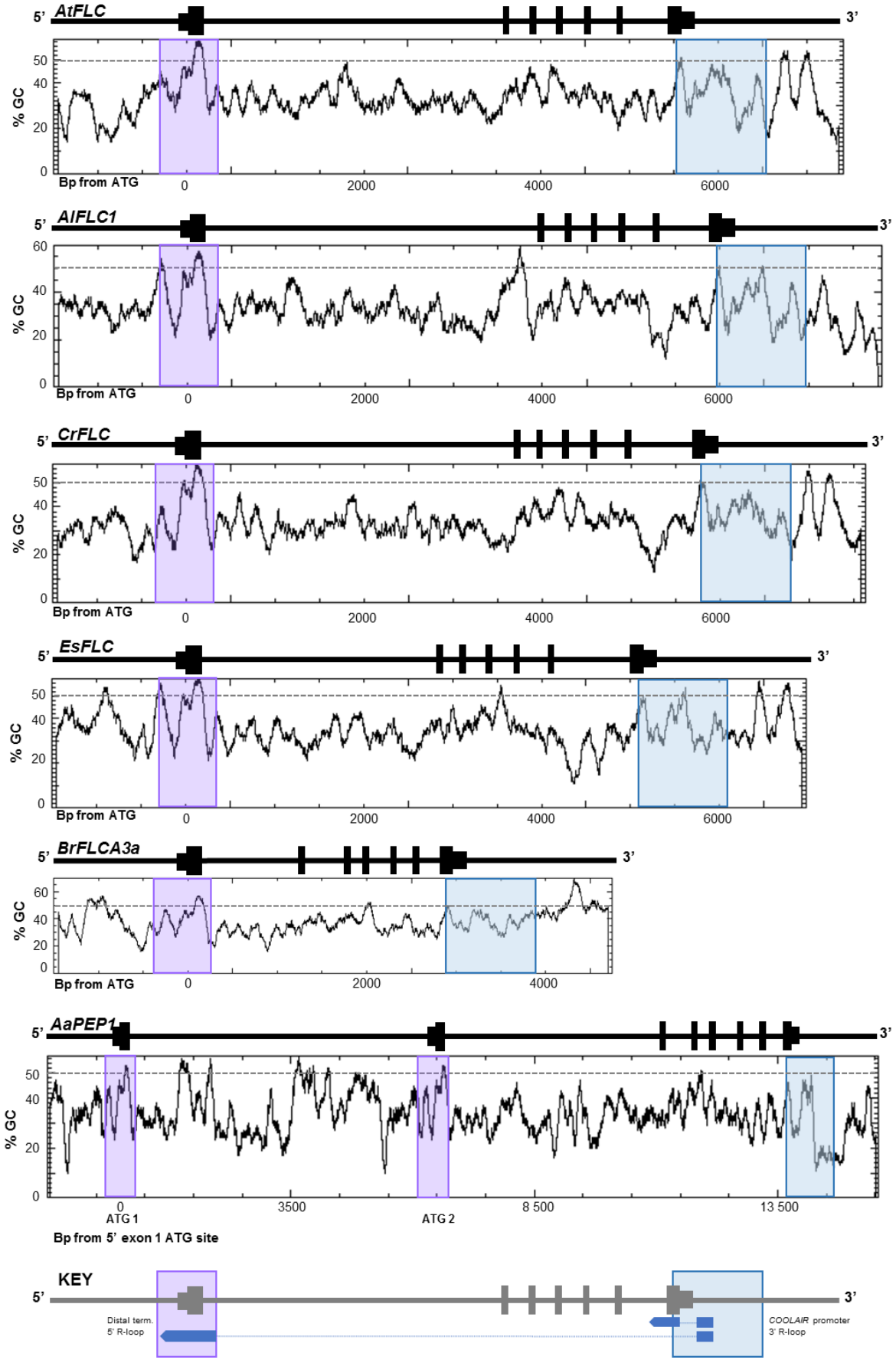


Figure 3.3: GC content of *FLC* gene across six Brassicaceae species. %GC content plot created using the Cpgplot tool provided by EMBOSS (http://www.ebi.ac.uk/Tools/seqstats/). Purple and blue boxes map *FLC* 5' and 3' regions, respectively. *FLC* sequences are from 1.5 kb upstream of the ATG start to 1.8 kb downstream of the translational stop.

3.2.2 **COOLAIR** *exons exhibit lower sequence conservation than protein-coding exons*

COOLAIR sequence diverges across the six analysed Brassicaceae species with evolutionary distance (Fig. 3.4A and Fig. 3. 5A). Sequence similarity is visibly lowest at regions that do not overlap *FLC* exons (blue boxes) or UTRs (dashed blue boxes) (Fig. 3.4B and Fig. 3.5B). The previously identified ~100 bp block of sequence homology within *COOLAIR* exon 1 (section 3.2.1) is less apparent from this multiple alignment analysis, although nucleotides 20-80 exhibit higher conservation than other non-overlapping regions.

It is useful to try to quantitatively determine whether *COOLAIR* exons are more, equally or less well-conserved than protein-coding exons and introns. Tables 3.1 to 3.7 give sequence identity for different regions of the *FLC* locus, based on multiple sequence alignments of the six species. Conservation across protein-coding exons is high, ranging from 86.9-95.4% sequence identity; whereas conservation across the full genomic locus (from 1.5 kb upstream of the translational start site to 1.8 kb downstream of the translational stop site) is much lower, ranging from 30.1-58.7% (Table 3.1; Table 3.2). This matches the high exonic and low intergenic/intronic sequence conservation predicted from the pair-wise sequence analyses given in Fig. 3.2. The distal *COOLAIR* exhibits higher conservation than the proximal, ranging between 68.6-87.8% sequence identity compared with 53.4-75.9%. Although this is lower than for protein-coding exons, it is considerably higher than for the full genomic sequence. As the full genomic sequence covers extensive regions up- and down-stream of *FLC,* and *COOLAIR* overlaps with protein-coding regions, a fairer analysis would be to compare *FLC* intronic regions with *COOLAIR* non-sense-exon regions. For this purpose, *FLC* introns 2-5 were selected as they do not contain any known sense or antisense regulatory regions and are in proximity to protein-coding exons [as is *COOLAIR*]. *FLC* introns 2-5 ranged from 41.1-84.6% similarity, with high sequence similarity in more closely related species and low sequence similarity in distantly related species. This is perhaps a fair representation of no selection over evolutionary time. *COOLAIR* proximal [without sense exon 7] exhibited sequence identities of 46.4-69.1%, slightly higher than across introns 2-5. In contrast, *COOLAIR* distal [without sense exon 1] exhibited sequence identities of 60.3-83.7%, often higher than intronic sequence similarity. The difficulty with these analyses is that *COOLAIR* exons also overlap *FLC* 5’ and 3’ untranslated regions and the cold-responsive nucleation region, likely to contain key regulatory elements, so it is difficult to determine whether selection here acts to preserve *COOLAIR* or other features.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*	*BrFLCA3a*
AtFLC		95.4	92.8	90.6	87.3	87.6
AlFLC1	95.4		93.6	91.0	87.3	87.1
CrFLC	92.8	93.6		89.8	86.6	86.8
EsFLC	90.6	91.0	89.8		88.1	91.1
AaPEP1	87.3	87.3	86.6	88.1		86.9
BrFLCA3a	87.6	87.1	86.8	86.9	86.9	

Table 3.1: % nucleotide identity of *FLC* mRNA from six species.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*b	*BrFLCA3a*
AtFLC		55.0	58.7	46.9	41.8	33.0
AlFLC1	55.0		47.9	37.2	37.4	31.0
CrFLC	58.7	47.9		48.5	42.0	34.6
EsFLC	46.9	37.2	48.5		40.9	37.9
*AaPEP1*b	41.8	37.4	42.0	40.9		30.1
BrFLCA3a	33.0	31.0	34.6	37.8	30.1	

Table 3.2: % nucleotide identity of full-length genomic *FLC* from six species.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*b	*BrFLCA3a*
AtFLC		84.6	57.0	41.5	42.1	41.9
AlFLC1	84.6		57.9	41.7	41.1	42.8
CrFLC	57.0	57.9		58.6	56.4	60.2
EsFLC	41.5	41.7	58.6		57.3	71.1
*AaPEP1*b	42.1	41.1	56.4	57.3		57.0
BrFLCA3a	41.9	42.8	60.2	71.1	57.0	

Table 3.3: % nucleotide identity of *FLC* introns 2-5 from six species.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*	*BrFLCA3a*
AtFLC		75.9	71.3	66.5	64.5	53.9
AlFLC1	75.9		67.3	61.5	59.7	54.7
CrFLC	71.3	67.3		62.3	62.7	53.4
EsFLC	66.5	61.5	62.3		70.6	55.6
AaPEP1	64.5	59.7	62.7	70.6		56.2
BrFLCA3a	53.9	54.7	53.4	55.6	56.2	

Table 3.4: % nucleotide identity of *COOLAIR* proximal Class I.i from six species.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*	*BrFLCA3a*
AtFLC		69.1	68.3	60.2	60.7	56.4
AlFLC1	69.1		60.7	53.7	54.4	48.5
CrFLC	68.3	60.7		55.5	58.9	47.5
EsFLC	60.2	53.7	55.5		62.2	51.6
AaPEP1	60.7	54.4	58.9	62.2		49.5
BrFLCA3a	46.4	48.5	47.5	51.6	49.5	

Table 3.5: % nucleotide identity of *COOLAIR* proximal I.i (without sense exon 7) from six species.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*	*BrFLCA3a*
AtFLC		87.8	82.1	79.2	71.4	73.1
AlFLC1	87.8		82.1	80.3	72.8	73.9
CrFLC	82.1	82.1		76.7	71.4	72.0
EsFLC	79.2	80.3	76.7		72.7	75.1
AaPEP1	71.4	72.8	71.4	72.7		68.8
BrFLCA3a	73.1	73.9	72.0	75.1	68.6	

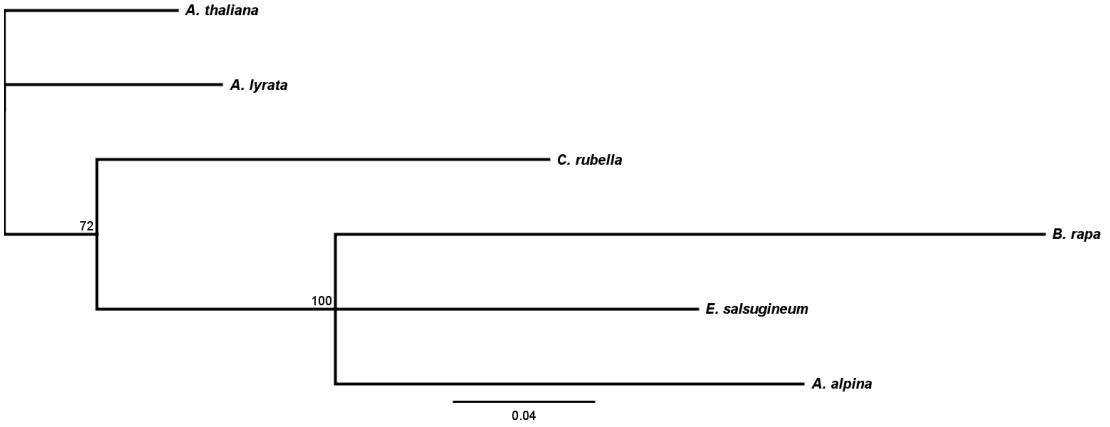
Table 3.6: % nucleotide identity of *COOLAIR* distal Class II.i from six species.

% identity	*AtFLC*	*AlFLC1*	*CrFLC*	*EsFLC*	*AaPEP1*	*BrFLCA3a*
AtFLC		83.7	75.9	73.0	64.1	65.2
AlFLC1	83.7		75.9	74.6	66.0	66.6
CrFLC	75.9	75.9		70.2	64.9	64.5
EsFLC	73.0	74.6	70.2		66.2	66.7
AaPEP1	64.1	66.0	63.9	66.2		60.3
BrFLCA3a	65.2	66.6	64.5	66.7	60.3	

Table 3.7: % nucleotide identity of *COOLAIR* distal Class II.i (without sense exon 1) from six species.

Tables 3.1-3.7. % nucleotide identity of *FLC* genomic regions from six species (*A. thaliana*, *A. lyrata*, *C. rubella*, *E. salsugineum*, *A. alpina* and *B. rapa*) based on Muscle multiple sequence alignment (Geneious R7). Full-length sequence is from 1.5 kb upstream of the translational start site to 1.8 kb downstream of the translational stop site.

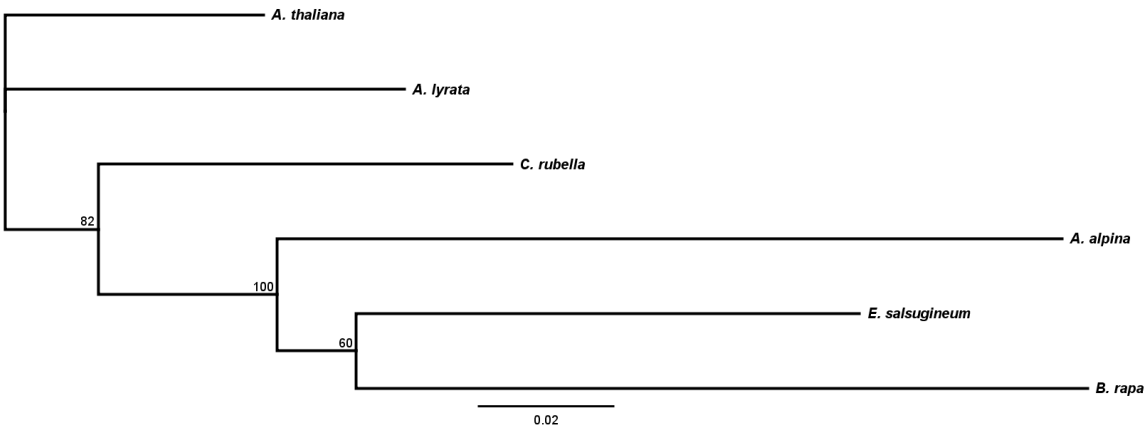
A Proximal I.i *COOLAIR* phylogenetic tree



B Proximal I.i *COOLAIR* sequence alignment

Figure 3.4: Sequence conservation of the proximal *COOLAIR* from six Brassicaceae species. (A) Phylogenetic tree created from MUSCLE multiple sequence alignment of *COOLAIR* proximal Class I.i transcripts using the HKY neighbour-joining model (Geneious R7). Topology tested with 1000 bootstrap replicates (consensus tree shown with % consensus support labelled). (B) ClustalW multiple sequence alignment of the same proximal transcripts from *A. thaliana* (AtFLC), *A.lyrata* (AlFLC1), *C. rubella* (CrFLC), *A. alpina* (AaPEP1), *E. salsugineum* (EsFLC) and *B. rapa* (BrFLC3). Alignment created with AlignX software, Vector NTI Advance 11.5, Invitrogen. Black/green represents the consensus residue derived from a single conservative residue at given position; blue/cyan, from a block of similar residues; red/yellow, from completely conservative residues; black/white, non-homologous residues.

A Distal II.i *COOLAIR* phylogenetic tree



B Distal II.i *COOLAIR* sequence alignment

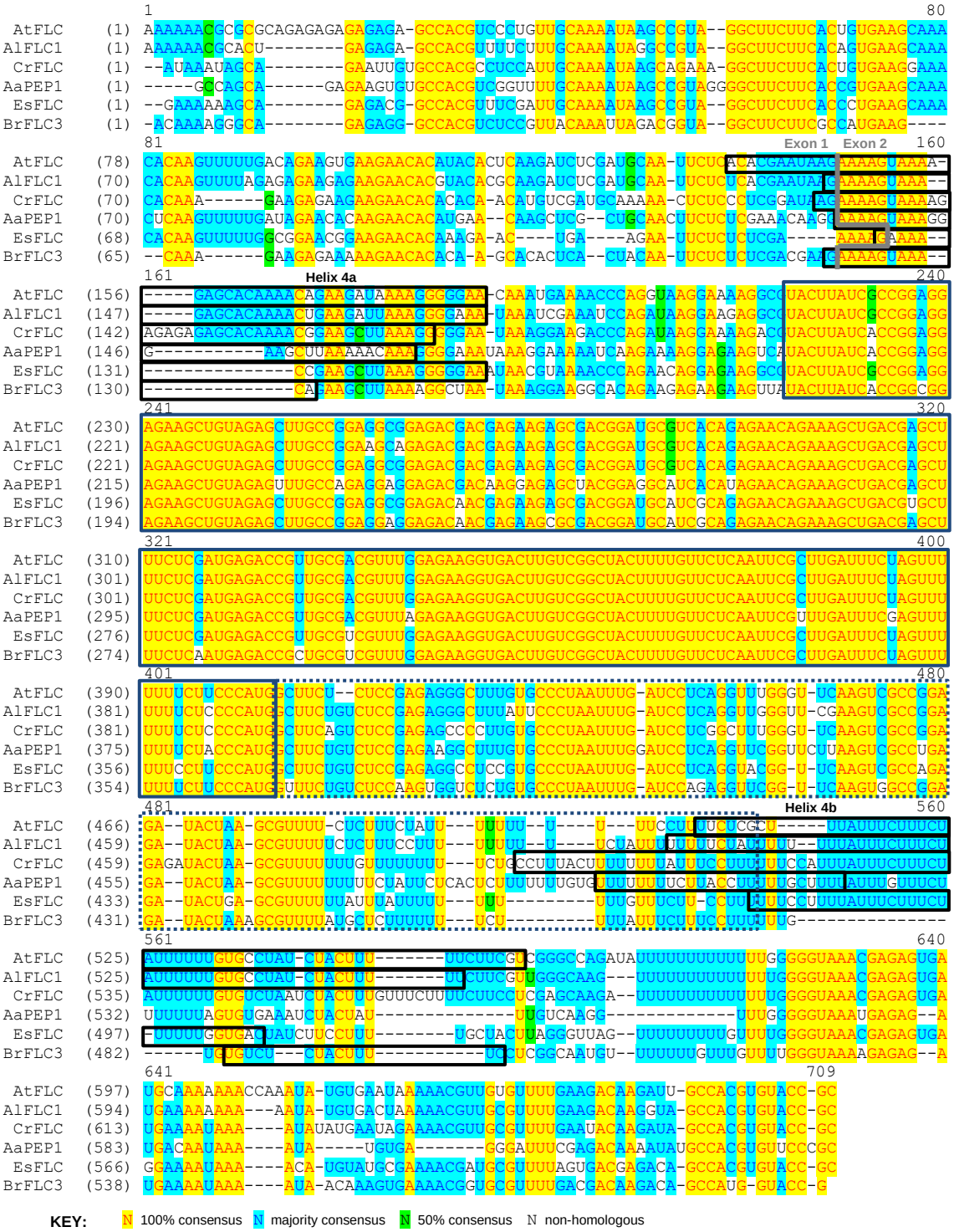
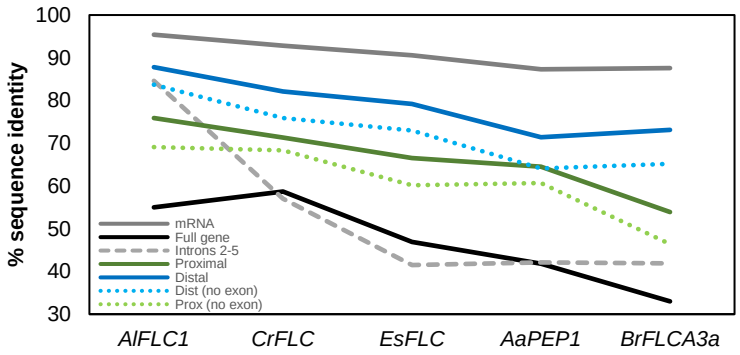


Figure 3.5: Sequence conservation of the distal *COOLAIR* across six Brassicaceae species (previous page). (A) Phylogenetic tree created from MUSCLE multiple sequence alignment of *COOLAIR* distal Class II.i transcripts using the HKY neighbour-joining model (Geneious R7). Topology was tested with 1000 bootstrap replicates (consensus tree shown with % consensus support labelled). (B) ClustalW multiple sequence alignment of the same distal transcripts from *A. thaliana* (AtFLC), *A.lyrata* (AlFLC1), *C. rubella* (CrFLC), *A. alpina* (AaPEP1), *E. salsugineum* (EsFLC) and *B. rapa* (BrFLC3). Alignment created with AlignX software, Vector NTI Advance 11.5, Invitrogen. Black/green represents the consensus residue derived from a single conservative residue at given position; blue/cyan, from a block of similar residues; red/yellow, from completely conservative residues; black/white, non-homologous residues. Helix 4a represents the left-hand strand and Helix 4b the right-hand stand of helix 4, which forms during folding (see Fig. 3.7); this consists of primarily non-sense-overlapping nucleotides. Adapted from Hawkes *et al*. (2016).

In summary, there has been strong divergence in the syntenic sequence underlying *COOLAIR* at the *FLC* locus across these six species. Regions of highest sequence similarity are found at the *FLC* protein-coding exons (Fig. 3.2). Sequence identities for different regions of *FLC* (for each species compared with *A.thaliana*) are presented graphically in Fig. 3.6A. This further demonstrates that proximal and distal *COOLAIR* conservation (represented by % sequence identity) is lower than that of protein-coding regions, but often higher than intronic regions. These relationships match evolutionary divergence of the six species (Fig. 3.6B).

A



B

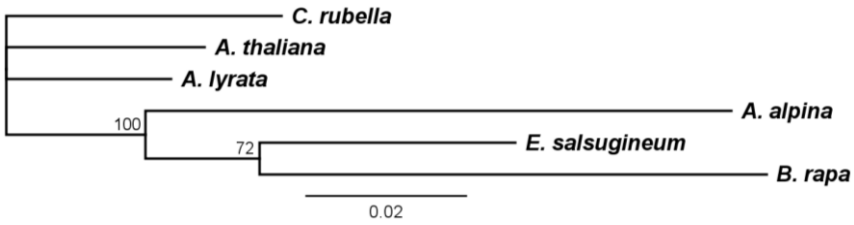


Figure 3.6: Summary of *FLC* and *COOLAIR* conservation across six Brassicaceae species. (A) % sequence identity of five species with *A. thaliana FLC*. Sequence identity data (presented in tables 3.1-3.7) is from MUSCLE multiple sequence alignment of genomic DNA plus the six other (labelled) segments (Geneious R7); dist = distal, prox = proximal. Five species are: *A. lyrata* (*AlFLC1*), *C. rubella* (*CrFLC*), *E. salsugineum* (*EsFLC*), *A. alpina* (*AaPEP1*) and *B. rapa* (*BrFLCA3a*). (B) Phylogenetic tree created from MUSCLE multiple sequence alignment of *FLC* mRNA using the HKY neighbour-joining model (Geneious R7). Topology was tested with 1000 bootstrap replicates (consensus tree shown with % consensus support labelled).

3.2.3 **Arabidopsis thaliana COOLAIR** *folds into elaborate secondary structures*

Another level of RNA conservation is secondary structure (Diederichs, 2014). Online tools to predict RNA structure are available; however, these often produce a long list of possible structures, none of which may be correct. Chemical probing *in vitro*, followed by *in vivo* confirmation, is a more effective way to validate secondary structure. We collaborated with the Sanbonmatsu group (Los Alamos National Laboratory) who use SHAPE and CMCT-probing to solve RNA secondary structure, supplemented by shotgun secondary structure (3S) determination (see Merino et al., 2005 for SHAPE-probing and Novikova *et al.*, 2013 for the 3S method). Chemical-probing revealed distal Class II.i and proximal Class I.i forms of *A.thaliana COOLAIR* fold into elaborate secondary structures *in vitro* (Fig. 3.7A and B; Hawkes *et al.*, 2016). Double-stranded helices consist of canonical Watson-Crick base pairs, but also non-canonical U-G and A-G base pairs. The latter can be found at the end of double helices, where they prevent further base stacking (Traub and Sussman, 1982). U-G 'wobble' pairs are a common feature in rRNA and tRNA folds, and can contribute to less or more stable structural features, depending on their position and structural context (Gautheret *et al.*, 1995).

The distal form is most interesting structurally, containing 12 helices (labelled H1-H12), seven stem loops, an intricate multi-way junction and two rare right-hand-turn (r-turn) motifs in the second exon (Fig. 3.7A; Hawkes *et al.*, 2016). r-turns are asymmetric 5' internal loops consisting of a large single-stranded region on the 5' side and a very short single-stranded region on the 3' side, with potential base pairing within the loop. These may be involved in RNA function, i.e. through protein or ligand binding (Montemayor *et al.*, 2014; Ren *et al*., 2016). The shared distal and proximal first exon is also highly structured with three helices (H1-H3) forming a three-way junction. The proximal second exon contains two additional regions of structure (H4-H6, capped with stem-loops), but is largely disordered with low reactivities indicating lack of structure (Fig. 3.7B). There is the potential for a pseudoknot to form between single-stranded sequence and the first internal loop of helix H4, which could lead to more complex tertiary structure (Fig. 3.7B; Hawkes *et al.*, 2016). Regions which overlap with sense protein-coding exons are highlighted by coloured boxes in Fig. 3.7. Sense exon 1 encompasses most of helices H7, H8 and H9 and the second r-turn in the distal structure, suggesting that this domain will be highly conserved in other species (due to high sequence conservation).

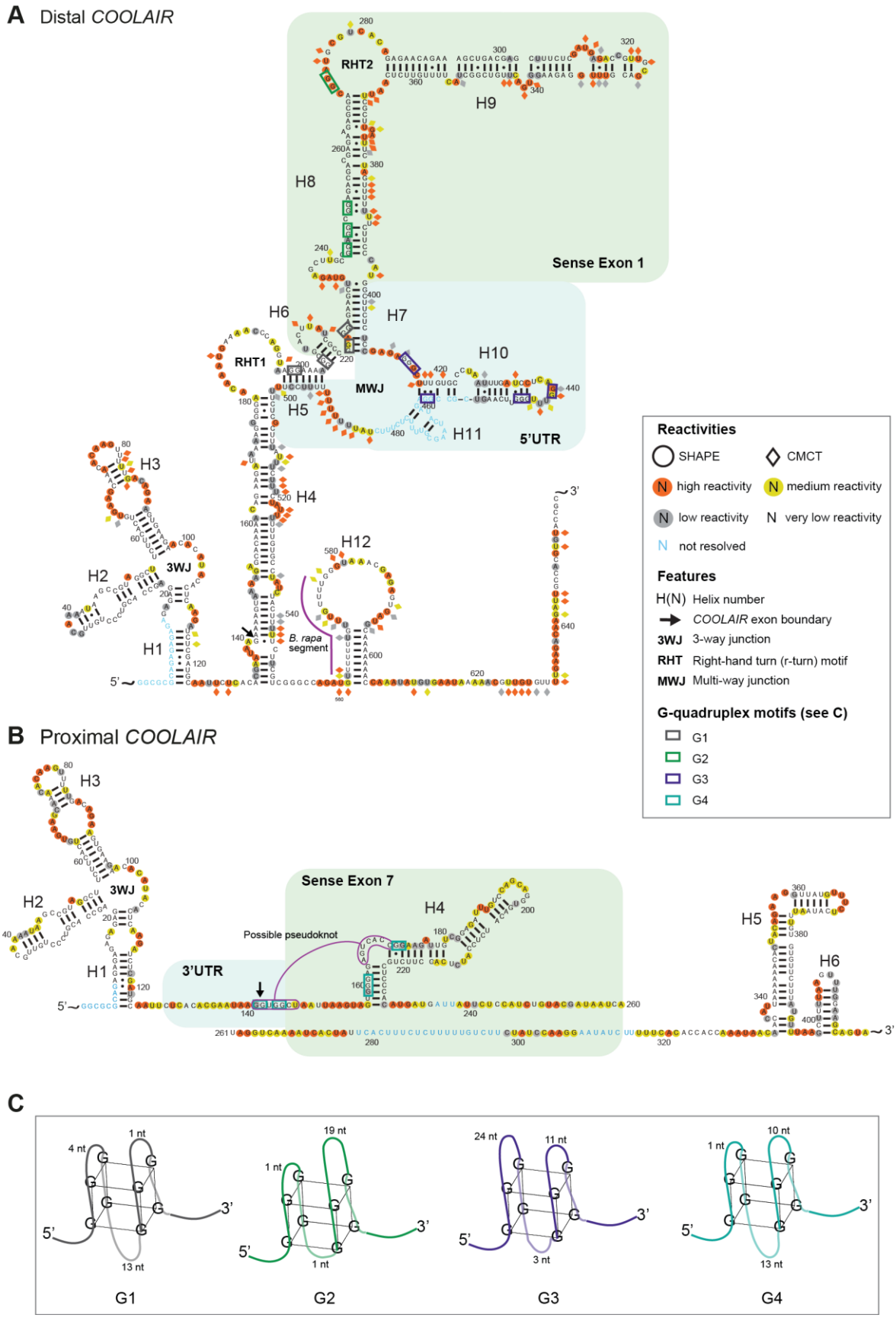


Figure 3.7: Secondary structures of the *A. thaliana* proximal and distal *COOLAIR* transcripts. (A) Secondary structure of the distal *COOLAIR* Class II.i from the *A. thaliana* Col-0 accession, based on SHAPE- and CMCT-probing experiments. Normalised SHAPE reactivity is represented as coloured circles and normalised CMCT reactivity as coloured diamonds. A short segment was replaced with *B. rapa* sequence to improve reactivity data read. (B) Secondary structure of proximal *COOLAIR* Class I.i from *A. thaliana* Col-0, based on SHAPE-probing experiments (as above). (C) Predicted G-quadruplex structures G1-G4; positions are mapped onto (A) and (B). Figure 3.7A and B reproduced from Hawkes *et al*. (2016).

The more variable 5'UTR encompasses the multi-way junction, whereas the extensive helix H4 consists of intronic sequence only. Sense exon 7 encompasses the helix H4 domain in the proximal structure, and therefore may be more highly conserved (based on our knowledge of sequence similarity alone) than the other structural features.

We have confidence in these secondary structures because nucleotides with high SHAPE reactivities are predominantly located in the terminal loops, internal loops and junction regions, and nucleotides with lower reactivities in the helices. There are instances where reactive nucleotides are thought to be base pairing and are assigned to a helix, such as in helices H8 and H9; this allocation is based on context (i.e. surrounding nucleotide reactivities). This is more common where the proposed base pairing nucleotide is adjacent to a single-stranded loop region, and so is potentially more reactive to the SHAPE reagent. This has been observed in other RNAs with validated secondary structure, i.e. the 16S rRNA (Noller and Woese, 1981). The SHAPE-determined secondary structures in Fig. 3.7 were verified by (1) repeat SHAPE-probing, (2) additional CMCT analysis and (3) the 3S method (see Fig. S1 for 3S validation of the distal folds).

G-quadruplexes are another structural feature that can form in RNA or DNA (Duquette et al., 2004; for reviews, see: Burge et al., 2006; Lipps and Rhodes, 2009). The formation of G-quadruplexes would form a competitive alternative to RNA base pairing under certain conditions. The *COOLAIR* distal and proximal sequences were further analysed for the presence of G-quadruplex forming motifs. Canonical G-quadruplex motifs were not identified, but non-canonical G-quadruplexes with longer and mixed length loops or 'bulge' nucleotides have been reported (Palumbo *et al.*, 2010; Mukundan and Phan, 2013; Jodoin *et al.*, 2014). Four possible non-canonical G_2 motifs (G1-G4) are presented in Fig. 3.7C, with the sequence from which they are derived highlighted within the SHAPE secondary structures in Fig. 3.7A and Fig. 3.7B. Alternative G_2 motifs could also form that combine some of the above, or use further afield G_2 sequences. These motifs are found in highly structured regions of both *COOLAIR* forms and could potentially disrupt normal folding under conditions that favour G-quadruplex formation. G4 overlaps with the potential pseudoknot in the proximal *COOLAIR*, providing an alternate tertiary structure. Further experimental evidence is needed to validate whether non-canonical G-quadruplexes form from *COOLAIR* sequence.

3.2.4 **cis** *polymorphisms can disrupt* **COOLAIR** *secondary structure in* **Arabidopsis** *accessions*

In the above experiments, we probed *COOLAIR* RNA from the widely-used *Arabidopsis* Columbia (Col-0) accession to determine secondary structure. Other functionally distinct *FLC* haplotype groups exist in *A. thaliana* accessions from different parts of the world. The Sanbonmatsu group carried out further chemical (SHAPE) probing on the distal *COOLAIR* isoform from two additional *Arabidopsis* accessions, Lov-1 and Var2-6, which contain *cis* polymorphisms at the *FLC* locus and within *COOLAIR*-derived regions. These three SHAPE-probed structures were then used to predict secondary structure in other *A. thaliana* accessions (Fig. 3.8). The five example accessions represent four of the five major *FLC* haplotype groups (Hap1, Hap3, Hap5, Hap11) from Li *et al.* (2014), in addition to Lov-1 from Hap9. The secondary structure of *COOLAIR* from the fifth major haplotype group (Hap13) is represented by Col-0 in Fig. 3.7. These also cover the vernalization response types characterised by Li *et al.* (2014), with the rapid vernalizers RV1-2 represented by Edi-0 (Hap3) and Col-0 (Hap13) and the slow vernalizers SV1-4 represented by Lov-1 (Hap9), Ull2-5 (Hap1), Bro1-6 (Hap5) and Var2-6 (Hap11). As there are only a small number of polymorphisms at *COOLAIR*, many of which are shared between more than one accession, we can be reasonably confident in the non-probed structure predictions.

The U to A SNP (with Col-0) in helix H3 is found in all five accessions and converts the first double helix to a larger terminal loop. This is an example of a RiboSNitch, a single nucleotide polymorphism which causes a significant change in secondary structure (Wan *et al.*, 2014; for review see: Lokody, 2014). However, this SNP is found in both rapid and slow vernalizers and has not been linked to changes in *COOLAIR* or *FLC* expression. The U nucleotide which gives rise to the internal loop and small terminal loop in H3 is in fact unique to Col-0. In evolutionary terms, it is more likely the *A. thaliana* distal *COOLAIR* isoform originally contained a large terminal loop at H3, as seen for the five accessions in Fig. 3.8A-E. The A-U SNP in Col-0 may have arisen later and favoured formation of an internal loop and small terminal loop structure, as seen in Fig 3.7A-B.

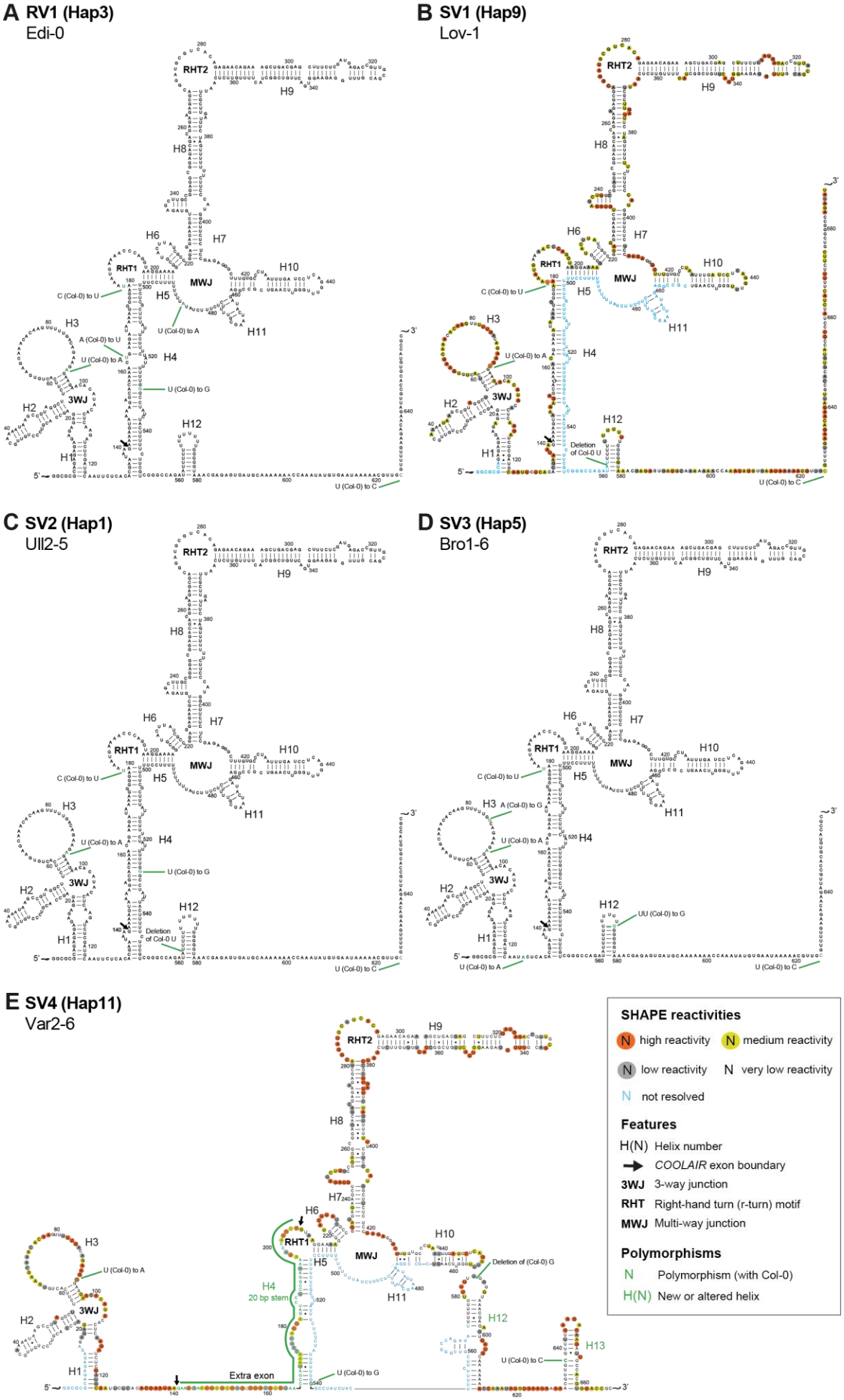

A RV1 (Hap3)
Edi-0
B SV1 (Hap9)
Lov-1
C SV2 (Hap1)
Ull2-5
D SV3 (Hap5)
Bro1-6
E SV4 (Hap11)
Var2-6
RHT2
RHT1
MWJ
3WJ
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
H11
H12
H13
C (Col-0) to U
A (Col-0) to U
U (Col-0) to A
U (Col-0) to G
A (Col-0) to G
U (Col-0) to C
U (Col-0) to U
UU (Col-0) to G
Deletion of Col-0 U
Deletion of (Col-0) G
20 bp stem
Extra exon
SHAPE reactivities
high reactivity
medium reactivity
low reactivity
very low reactivity
not resolved
Features
H(N) Helix number
COOLAIR exon boundary
3WJ 3-way junction
RHT Right-hand turn (r-turn) motif
MWJ Multi-way junction
Polymorphisms
N Polymorphism (with Col-0)
H(N) New or altered helix

Figure 3.8: Secondary structure of *COOLAIR* distal transcript from five *A. thaliana* accessions, representing major *FLC* haplotype groups (previous page). Predicted (A, C, D) and chemically-probed (B, E) secondary structures of the distal *COOLAIR* Class II.i transcript from *A. thaliana* Edi-0 (A), Lov-1 (B), Ull2-5 (C), Bro1-6 (D) and Var2-6 (E). These represent five major *FLC* haplotype groups and vernalization response types (labelled) (Li *et al*., 2014). Chemical-probing was carried out by the Sanbonmatsu group; normalised SHAPE reactivity is represented by coloured circles and normalised CMCT reactivity by coloured diamonds. Figure 3.8E is reproduced from Hawkes *et al*. (2016).

Aside from the H3 RiboSNitch, five out of six haplotype groups (including Col-0, haplotype 13) exhibit the same or near-identical secondary structures for the distal *COOLAIR* (Fig. 3.8). Identical secondary structures across different accessions within a species could occur when (1) there are no sequence polymorphisms, (2) sequence polymorphisms are in or adjacent to single-stranded or loop regions, or (3) where sequence polymorphisms occur in double-stranded helices but do not significantly alter structure. Covariant base pair mutations could also give rise to identical secondary structures; however, shorter evolutionary distances within (as opposed to between) species make these less likely to arise. Here, sequence polymorphisms are primarily located within or adjacent to single-stranded or loop regions. This could indicate purifying selection of the sequence in structure-forming regions, selecting against mutations which disrupt a double-stranded helix. Alternatively, this could be a chance correlation given the low number of polymorphisms (4-6) between accessions.

Analysis reveals that > 75% of SNPs (between each accession and Col-0, or between Col-0 with all accessions) cause minor or no disruption to actual or predicted secondary structure (Table 3.8). This includes all SNPs within *COOLAIR* exons, but excludes the change in splicing that gives rise to the altered Var2-6 secondary structure, which is discussed below. The disruptive RiboSNitch in helix H3 is found only in Col-0; accordingly this SNP is allocated to Col-0, whereas the other SNPs are all recorded as if Col-0 is the canonical sequence.

Species	Accession	SNPs	No effect	Disruption	% conserving
A. thaliana	Col-0	10*	9	1	90%
	Lov-1	3	3	0	100%
	Edi-0	5	4	1	80%
	Ull2-5	4	3	1	75%
	Bro1-6	6	6	0	100%
	Var2-6	4*	4	0	100%

Table 3.8: Effect of *Arabidopsis thaliana* polymorphisms on distal Class II.i secondary structure. SNPs are recorded for Col-0 against all other accessions (highlighted in grey), where all but one are considered canonical and therefore automatically non-disruptive. SNPs are then recorded for each accession against Col-0. *This excludes the intronic SNP that alters distal splicing in Var2-6.

The lack of structural change appears to be primarily because the above SNPs occur in or adjacent to single-stranded non-structural or looped region. These include the A to U SNP enlarging an internal loop in H4, the C to U SNP in the first r-turn, the deletion of a U or a G in H12 (enlarging or shortening the loop or helix), and the U to C SNP in the single-stranded region at the 3' end. Ull2-5 and Edi-0 contain a U to G SNP which could prevent base pair formation in H4, and has therefore been classified as disruptive to secondary structure. It is considered unlikely, however, that significant disruption of H4 will occur as this is a single base pair break within a long helix; H4 has other internal loops which do not appear to disrupt its stability. It is probable that the four base pairs beneath and five base pairs above this break will remain paired, as there are other stable helices of these lengths or less (down to two base pairs) in the SHAPE-probed Col-0, Lov-1 and Var2-6 structures, i.e. in helices H1, H2, H3, H4, H6, H10, H11. In addition, we know that H4 is maintained in Var2-6, despite a much larger loss of base pairing and resultant internal loop, suggesting this helix is persistent and stable. An alternative possibility is that a non-canonical A-G base pair could form, as found in the Col-0 distal helix H4.

The predominance (> 75%) of non-disruptive SNPs could be a consequence of purifying selection (selection against disruptive SNPs) over evolutionary time. The Var2-6 structure, on the other hand, suggests positive selection of structural changes where these confer an advantage under specific environmental conditions. Together, these data may indicate a low level of purifying selection at *COOLAIR*, with positive selection where major changes confer advantage. Alternatively, it may be that structure is typically retained even with a larger number of polymorphisms, with very few nucleotides being potential RiboSNitches (so unable to confer major structural alterations). Perhaps a few SNPs within a double-stranded helix are manageable, but there is a breaking point where structure becomes significantly disrupted.

Secondary structure is substantially altered in haplotype group 11, represented by Var2-6 (Fig. 3.8E; Hawkes *et al.*, 2016). Var2-6 is a late flowering accession from northern Sweden; 46 accessions containing the Var2-6 *FLC* haplotype flowered significantly later than 101 accessions carrying other *FLC* haplotypes under non-vernalized conditions, with 32/46 not having flowering by 120 days (Li *et al.*, 2015). High *FLC* expression was mapped to a *cis* SNP which prevents canonical splicing of the *COOLAIR* distal II.i transcript (Li *et al.*, 2015). This induces a shift to a downstream distal splice acceptor site, shortening the terminal exon and encouraging inclusion of an internal exon (the distal Class II.iv transcript in Fig. 3.1). Consequences of the altered splicing are (1) a shorter and

potentially less stable helix H4 (17 versus 37 bp in Col-0 with a 14 bp reactive internal loop); (2) a longer unstructured linker region between H1 and H4; and (3) a more structured 3' end with a bifurcated H12 and additional H13. Any or all of these altered features could contribute to functional differences in *COOLAIR* between Var2-6 and Col-0. The change in helix H4 is particularly interesting considering the associated change in position of the first r-turn.

What is perhaps most surprising is that there was not a more dramatic change in global secondary structure, given the dramatic change in underlying sequence: the replacement of sequence pertaining to the left-hand side of helix H4 with sequence from part of *FLC* intron 1 (the additional exon). The addition of entirely new sequence which is still (at least in part) able to base pair with the 3' *COOLAIR* sequence that forms the right-hand side of helix H4 is impressive. This could be in consequence of this sequence mutating rapidly to allow base pairing and H4 formation, or targeted splicing of this internal exon because it was already capable of partial restoration of H4. Given the lower rate of polymorphisms between *Arabidopsis* accessions than between species the latter seems more likely. Alternatively, there may be potential for multiple random stretches of sequence to base pair with this region and partially maintain H4 in order to maintain an energetically-favourable global structure. Whichever the case, these data support the need for H4 and the first r-turn to be maintained, without which the global structure would collapse.

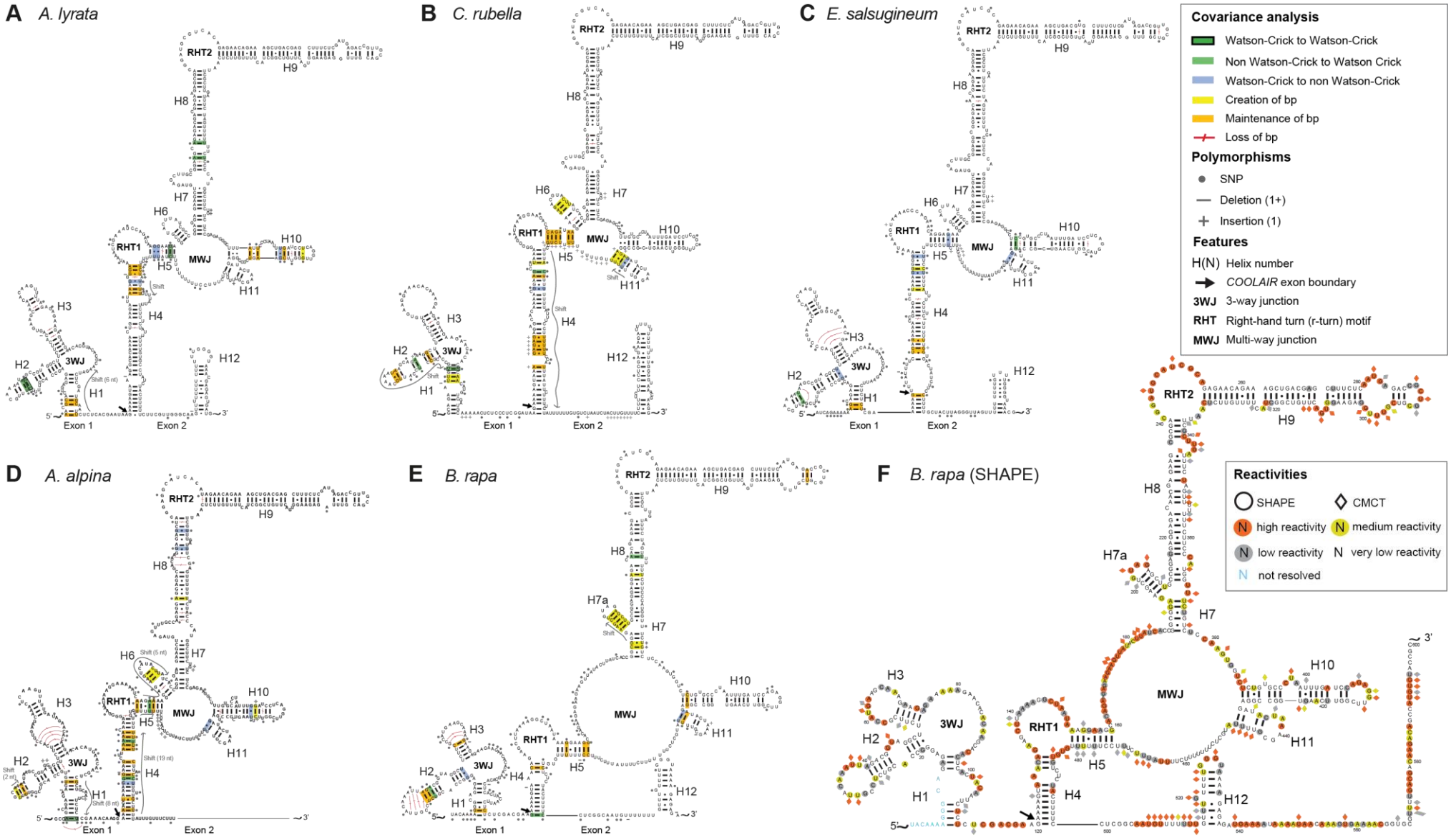
Within *Arabidopsis thaliana* significant changes in *COOLAIR* secondary structure are therefore associated with large changes in sequence, rather than with isolated SNPs within a structural domain. In the Var2-6 example there is correlation with flowering time, supporting the idea that *COOLAIR* secondary structure is functionally-relevant.

3.2.5 Global COOLAIR *structure is conserved across six Brassicaceae species*

For protein-coding mRNAs, conservation of amino acid codons is maintained by purifying (negative) selection. Non-coding RNAs do not necessarily need to maintain primary nucleotide sequence. If secondary structure is important for function this may instead be constrained, for example, through covariant base pair mutations. We predicted secondary structures of the distal *COOLAIR* transcript for five species within the family Brassicaceae, *A. lyrata, A. alpina, C. rubella, E. salsugineum* and *B. rapa*, representing ca. 13-43 million years divergence from *A. thaliana* (Koch and Kiefer, 2005; Beilstein *et al.*, 2010). In contrast to the *Arabidopsis* accessions analysed above, these contain a relatively high number of polymorphisms at *FLC* and *COOLAIR*, which increase with evolutionary

divergence (sections 3.2.1 and 3.2.2). We predict that with a certain degree of sequence dissimilarity global secondary structure will collapse *unless* other methods of conservation, such as base pair covariance, kick-in. Structure predictions were made using the manual strategy developed by Weinberg *et al.* (2007) of initial sequence alignment matching syntenic stretches of higher sequence identity, followed by re-alignment of lower sequence identity regions according to structural domains from a SHAPE-probed structure (in this case, that of *A. thaliana* Col-0). For example, *COOLAIR* helices H8-H9, which are antisense to a highly-conserved protein-coding region of *FLC* (containing the MADS box motif), were used to align homologous sequences across the five species. Next, stretches of sequence flanking H8-H9 were shifted to improve alignment with helices in *A. thaliana*. This was repeated and iterated outwards towards the 5' and 3' ends.

Predictions revealed strong homology of the distal *COOLAIR* global secondary structure, with major structural elements conserved (Fig. 3.9). The 3-way junction, the two r-turns, the multi-way junction, and the terminal region of H4 are conserved across all six species. Five species contain a five-way central junction, whilst one species (*B. rapa*) contains a four-way junction due to loss of helix H6. Helices H7, H8, H9, H10 and H11 are conserved across all six species but exhibit some length variation. There is strong sequence conservation underlying helices H7-H9 due to their overlap with the protein-coding sense exon 1. This means that the second r-turn is always conserved and there is minimal opportunity for base pair covariance in these regions. Covariant base pair flips (green and blue boxes) are found in all helices apart from H3, supporting their structural conservation. We also observe maintenance of base pairs (orange boxes, where base pairing at a given position is maintained through insertions or shifts in the underlying sequence) and the creation of new base pairs (yellow boxes, where polymorphisms induce base pairing at a new position). Shifts in helix-forming sequence (grey arrows) are predicted at helix H4 in both *C. rubella* and *A. alpina*, and are critical for maintaining the first r-turn and consequently global secondary structure. Loss of base pairs (red lines) primarily occurs adjacent to a single-stranded loop, hence the loop is enlarged but the sequence not dramatically altered, i.e. the *A. alpina* H3 internal loop (Fig. 3.9D) and the *B. rapa* H3 terminal loop (Fig. 3.9E). In addition, we observe overall greater sequence variation in single stranded than double-stranded regions (SNPs marked by grey dots). In summary, a combination of factors support conservation of global secondary structure: (1) sequence conservation, (2) covariant base pair mutations, (3) sequence shifts, (4) insertions and (5) reduced polymorphisms in double-stranded versus single-stranded regions.


A A. lyrata
B C. rubella
C E. salsugineum
D A. alpina
E B. rapa
F B. rapa (SHAPE)
RHT2
RHT1
MWJ
3WJ
H1
H2
H3
H4
H5
H6
H7
H7a
H8
H9
H10
H11
H12
Exon 1
Exon 2
Covariance analysis
Watson-Crick to Watson-Crick
Non Watson-Crick to Watson Crick
Watson-Crick to non Watson-Crick
Creation of bp
Maintenance of bp
Loss of bp
Polymorphisms
SNP
Deletion (1+)
Insertion (1)
Features
H(N) Helix number
COOLAIR exon boundary
3WJ 3-way junction
RHT Right-hand turn (r-turn) motif
MWJ Multi-way junction
Reactivities
SHAPE
CMCT
high reactivity
medium reactivity
low reactivity
very low reactivity
not resolved

F**igure 3.9: Secondary structure of the *COOLAIR* distal transcript across five Brassicaceae species (previous page).** Predicted (A-E) and chemically-probed (F) secondary structures of the distal *COOLAIR* Class II.i transcript from *A. lyrata* (A), *C. rubella* (B), *E. salsugineum* (C), *A. alpina* (D) and *B. rapa* (E and F). Covariance and polymorphism analyses were performed via sequence and structure comparison with the *A. thaliana* Col-0 distal Class II.i transcript. Chemical-probing of *B. rapa* (F) was carried out by the Sanbonmatsu group; normalised SHAPE reactivity is represented by coloured circles and normalised CMCT reactivity by coloured diamonds. Figure adapted from Hawkes *et al*. (2016).

Although global secondary structure is preserved, there are differences in structural elements in the more evolutionarily distinct species. Areas where we see the greatest variations are in the length and presence of helix H12 at the 3' end, and in the length and stability of helix H4 – it has a large internal loop in *E. salsugineum* and is considerably shorter in *A. alpina* and *B. rapa*.

As the predicted secondary structure of *B. rapa COOLAIR* was most divergent, the Sanbonmatsu lab performed SHAPE and CMCT analysis to confirm folding (Fig. 3.9F; Hawkes *et al.*, 2016). This Class II.i transcript is spliced in the same way as the *A. thaliana* Col-0 isoform, but contains multiple polymorphisms, including a deletion in the region underlying the right-hand side of helix H4. We know from Var2-6 that changes in sequence can significantly alter secondary structure, but covariance analysis predicted the global *B. rapa* structure would be maintained. Chemical-probing confirmed that global structure was conserved *in vitro*, with covariant base pair flips in the helices and greater variation in the single stranded regions. Strong sequence conservation of the protein-coding exon (H8 and H9) retains the second r- turn. Similar to the *A. thaliana* Var2-6 isoform and the *A. alpina* transcripts, H4 was significantly shorter, partially due to an 11 bp deletion disrupting its 5' side. The 17 bp stem of H4 in Var2-6 could be responsible for its altered behaviour and late flowering phenotype; the *B. rapa* distal *COOLAIR*, with its even shorter stem, may therefore behave more similarly to Var2-6 than Col-0. *B. rapa* genotypes exhibit a wide range of morphological and flowering phenotypes; it is interesting to speculate that this could, in part, be a consequence of sequence polymorphisms between *COOLAIR* transcripts. Maintenance of even a short H4, primarily through a short 7 bp stretch of sequence conservation, preserves the first r-turn connecting H4 and H5. The multi-way junction is present but contains one less helix (H6) relative to the other species. In addition, H1 and H7 are less stable than for other species, and some helices have shifted or changed in length. H3 has a large terminal loop and no internal loop, reminiscent of the Var2-6 structure. Potential base pairing between the loop of the multi-way junction (which contains 13 nt with low SHAPE reactivities) and the nucleotides forming the 3' side of H12 could impact tertiary folding.

3.2.6 Transcription of COOLAIR *from syntenic regions of* FLC

The next step was to determine transcriptional conservation of *COOLAIR* at the *FLC* locus. Primers designed from the predicted secondary structures in Fig. 3.9 were used to confirm the *in vivo* presence of the proximal and distal *COOLAIR* isoforms in *A. lyrata, C. rubella* and *B. rapa* (Fig. 3.10A). Nested touch-down PCRs were used to detect a specific product (rather than normal PCR) due to the low expression of *COOLAIR* and the presence of multiple *FLC* copies in *A. lyrata* and *B. rapa*. Three major splice variants were identified and classified according to the established *A. thaliana* classes in Fig. 3.1: the proximal Class I.i and the distal Class II.i and II.ii transcripts. Splice sites are largely conserved, with the exception of the proximal intron 3' acceptor splice site in *B. rapa*, and the distal Class II.ii large intron 3' acceptor site in *C. rubella* (Fig. 3.10B). The proximal Class I.ii unspliced transcript also appears to be present; this may alternatively represent the unspliced distal transcript or (most likely) both. There is some ambiguity about the origin of this transcript. This is because occasionally these nested PCRs also amplify the sense transcript, despite the use of gene-specific *COOLAIR* reverse primers in the RT (reverse transcription) reaction. These only bind to and amplify the antisense transcripts at *FLC*; however, the reaction can function at a low level without any added primers – potentially because of random priming from remaining fragments of non-degraded DNA (Haddad *et al*., 2007) – hence, in subsequent nested PCRs it can be possible to amplify from a low level of sense-strand specific transcripts.

The same proximal Class I.i isoform is conserved in all four species (including *A. thaliana*). In contrast, two different distal isoforms were identified. Differential distal splicing in *A. thaliana* accessions (Var2-6 vs. Col-0) alters *FLC* expression (Li *et al*., 2015) and may be equally important across species. Comparison of these loci is complicated by ancient polyploidization and tandem duplication events creating multiple copies of *FLC* in *B. rapa* and *A. lyrata*. One of four *B. rapa* copies (*FLCA3a*) and one of two *A. lyrata* copies (*FLC1*) were analysed. As loci diverge independently over time, it may be that each expresses unique splicing isoforms. Indeed, a distal *COOLAIR* isoform with an alternate 3' acceptor site was recently identified at the *FLCA2* locus in *B. rapa* (Li *et al*., 2016). Detection of *COOLAIR* isoforms with similar architecture to *A. thaliana* supports the predicted conservation of secondary structure.

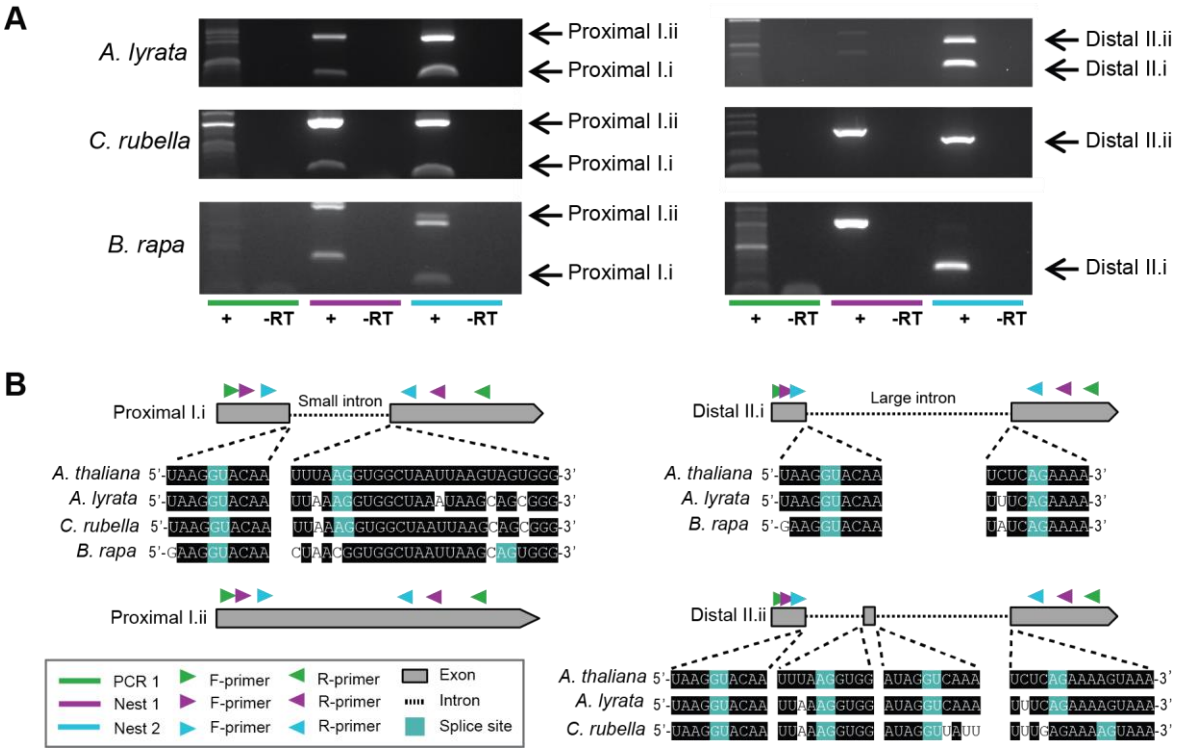


Figure 3.10: *COOLAIR* proximal and distal transcript architecture across four Brassicaceae species. (A) RT-PCR experiments probing for the proximal (left) and distal (right) forms of *COOLAIR* from non-vernalized *A. lyrata*, *C. rubella* and *B. rapa* leaf tissue. Initial RT-PCR (green line) was followed by two nested PCRs (purple and blue lines), where the + column is the cDNA sample and the -RT column is the DNA contamination (no reverse transcriptase) control. Different splice variants are labelled according to the *A. thaliana* classes in Fig. 3.1. (B) Sequencing RT-PCR nest 2 products revealed the major *COOLAIR* isoforms, with grey boxes in the schematic representing exon positions and triangles representing primer positions. Sequences were aligned with *A. thaliana COOLAIR* Col-0 to compare splice sites, highlighted in blue. Figure reproduced from Hawkes *et al*. (2016).

3.3 Discussion

Long non-coding (lnc)RNAs are often less well-conserved than their protein-coding counterparts (Wang *et al.*, 2004; Marques and Ponting, 2009). Critics suggest that without evolutionary conservation, we cannot have biological function (Wang *et al.*, 2004; Struhl, 2007; Graur *et al.*, 2013). Although the conservation of lncRNAs, as a group, is greater than expected by chance (i.e. by neutral evolution), it has been suggested that whilst some are biologically meaningful, the majority represent transcriptional noise (Struhl, 2007). The problem with such conclusions is their focus on primary nucleotide sequence conservation. lncRNAs that function through interactions with 3D structure elements, for example, can exhibit low conservation of primary nucleotide sequence but remain biologically relevant. We propose that the *Arabidopsis thaliana COOLAIR* transcripts are one such lncRNA.

Primary sequence conservation at *COOLAIR* is lower than for protein-coding sense exons at *FLC*, with sequence similarity decreasing with evolutionary divergence. This is in agreement with other studies that consider conservation of lncRNAs (Marques and Ponting, 2009; Kutter *et al.*, 2012). Comparison with non-regulatory intronic regions (introns 2-5) in *FLC*, however, suggests a low level of purifying selection; sequence similarity is higher across species for *COOLAIR* than for *FLC* intron sequences, except for between the closely related *A. thaliana* and *A. lyrata*. Assuming introns 2-5 are subject to normal evolutionary drift, then *COOLAIR* appears to be subject to some selective pressure. Selective pressure is not evident in *A. lyrata*, which contains the lowest number of *COOLAIR* polymorphisms. This suggests SNPs can be tolerated, up to a point. The difficulty with this kind of analysis at *COOLAIR* is the overlap with the protein-coding sense exons of *FLC* which are subject to high selective pressure. This was resolved somewhat by analysis of non-overlapping regions only; however, these still overlap with the sense promoter, 5'UTR, 3'UTR and nucleation region, some of which contain critical sense-regulatory elements. We therefore cannot be confident that there is selective pressure to preserve *COOLAIR* exon sequence. *COOLAIR* exon 1 contains a ~100 nt conserved block; conservation in this region was also observed by Castaings *et al.* (2014). Within this there is a conserved GC peak, first observed by Sun *et al.* (2013), underlying the *A. thaliana* 3' R-loop. Higher sequence restraint in this region may therefore act to conserve transcription of *COOLAIR* from its promoter (which extends downstream of exon 1), as well as the presence of a regulatory R-loop. The GC peak appears in all species, including *B. rapa*, for which sequence conservation in this region is less convincing. These regions should be analysed further to see if the 3'R-loop is indeed conserved. Presence of a highly-

conserved GC peak at the 5' end underlies another R-loop detected in *A. thaliana* (Zhe Wu, Dean lab, unpublished). Stability of the 3' R-loop moderates *COOLAIR* antisense activity; it is intriguing to speculate that the 5' R-loop similarly regulates the sense transcript, with distal *COOLAIR* transcription promoting R-loop formation and consequently sense downregulation.

Structural architecture is key to a mechanistic understanding of many functional RNAs. Secondary structure can be conserved without primary nucleotide conservation via covariant base pair mutations. To further investigate *COOLAIR* conservation and function, we determined the secondary structure of the *COOLAIR* proximal and distal transcripts using chemical probing experiments. These were performed *in vitro*; *in vivo* chemical probing will be useful to validate these structures and provide further insight into tertiary-folding or protein-binding within cells (Ding *et al.*, 2014). *In vitro* probing is, arguably, a critical first step in determining structure, given the greater difficulty of interpreting non-reactive bases *in vivo* – they could be involved in secondary base pairing, in tertiary folding or in protein-binding. We have confidence in the presented secondary structures, given that many RNA structures previously determined *in vitro* by chemical-probing have later been confirmed *in vivo* by crystallographic and binding studies (Noller and Woese, 1981; Roth and Breaker, 2009; Huet *et al.*, 2014). Recently, the crystal structure of the Pistol ribozyme validated and improved characterisation of the pseudoknot predicted from consensus secondary structure models (Harris *et al.*, 2015; Nguyen *et al.*, 2017).

Conservation of global *COOLAIR* distal secondary structure across different accessions and species, despite low sequence similarity in non-protein coding regions, supports its functional role. In effect, nature has maintained these structural features even from sequence of largely, or [in the case of the Var2-6 helix H4] completely, different composition. The mechanisms by which this was possible include covariant base pair mutations. Recently, the R-scape tool was developed to test for statistically significant covariation in RNA structure (Rivas *et al.*, 2016). R-scape analysis suggests covariant base pairs are statistically significant in the 5S ribosomal RNA and tRNA structures, but not in the *HOTAIR*, *SRA* or *Xist* lncRNA structures. R-scape found sequence variation in the latter were more frequently inconsistent than consistent with the proposed secondary structures. Here, we do not present a consensus structure, rather we have determined covariation by comparison of structure and sequence alignment for each predicted structure individually. This prevents any misinterpretation of consensus structures i.e. it is common practice to annotate a base pair as covarying if covariance is found in any single species in

consensus structures. It furthers allows a more comprehensive understanding of how global structure has been maintained, and to consider whether this is due to evolutionary selection or chance alone. In addition to base pair covariance, we see a predominance of SNPs in or adjacent to loop regions (i.e. the *Arabidopsis* accessions), conservation of short critical stretches of sequence (i.e. *B. rapa* H4), and shifting sequence from single-stranded regions to helices (i.e. *C. rubella* and *A. alpina* H4). The latter example of shifting sequence is particularly interesting because it suggests that when polymorphisms prevent canonical base pairing, one solution is to form the same structural features from different sequence regions if possible, thus retaining an energetically-favourable global structure. This is also seen in the more extreme example of Var2-6, where entirely new sequence from the internal exon is used to maintain H4. Sequence combinations within some helices makes such an occurrence more likely. For example, the right-hand side of H4 contains multiples stretches of U nucleotides, which can form Watson-Crick base pairs with A or non-Watson-Crick base pairs with G nucleotides, thus increasing the number of possible sequences it can base pair with.

Arabidopsis thaliana COOLAIR transcripts were found to be highly modular and organised by exon with three main structural domains: the 5', the central, and the stalk (Fig. 3.11A). This is comparable to the modularity observed in the mammalian *SRA*, *HOTAIR* and *Braveheart* lncRNAs (Novikova *et al.*, 2012; Somarowthu *et al.*, 2015; Xue *et al.*, 2016). This contradicts the argument that poorly conserved RNAs (in terms of nucleotide sequence) will lack functional relevance and therefore significant structural organization (Struhl, 2007). The distal *COOLAIR* transcript folds into an intricate secondary structure with a three-way junction, two r-turns and central multi-way junction. The three-way junction structural domain is shared with the proximal transcript. The proximal second exon is more disordered than that of the distal, but does contain two smaller structural domains. Distinct modular domains could have evolved to perform different functions. The structural architecture of Type II riboswitch RNAs, for example, facilitate spatially distinct binding sites (Montange and Batey, 2008). The proximal transcripts are functional in the autonomous floral promotion pathway, which downregulates *FLC* expression. The unique proximal H4-H6 structural domains could be involved in this mechanism. This is supported by conservation of H4-H6 in predicted secondary structures from four other species, including the possible pseudoknot interaction in three of the four (Hawkes *et al.*, 2016). A pseudoknot fold in the env25 Pistol ribozyme stabilises its tertiary structure and facilitates self-cleavage (Ren *et al.*, 2016). An alternate structural feature, a non-canonical G-quadruplex, could form from the same sequence, potentially competing with (or even

stabilising) the pseudoknot. Both pseudoknot and G-quadruplex structural elements can act as recognition sites for proteins or other RNAs. A G-quadruplex in a TERRA (telomeric repeat-containing) lncRNA was shown to bind the LSD1 complex, possibly inhibiting LSD1-catalyzed demethylation (Hirschi *et al.*, 2016). Given the epigenetic silencing of *FLC* chromatin during the cold, it would be interesting to investigate further the presence of G-quadruplexes and other (non-hairpin loop) structural elements at sense and antisense RNA or DNA. The 3' R-loop forms over the *COOLAIR* promoter, exon 1 and proximal exon 2, and is involved in regulating *COOLAIR* expression (Sun *et al.*, 2013). H1-H3 within the first exon of both proximal and distal transcripts may be involved in this mechanism. Higher conservation of the first *COOLAIR* exon (the ~100 nt block) [than of surrounding sequence] and of the three-way junction supports a functional role.

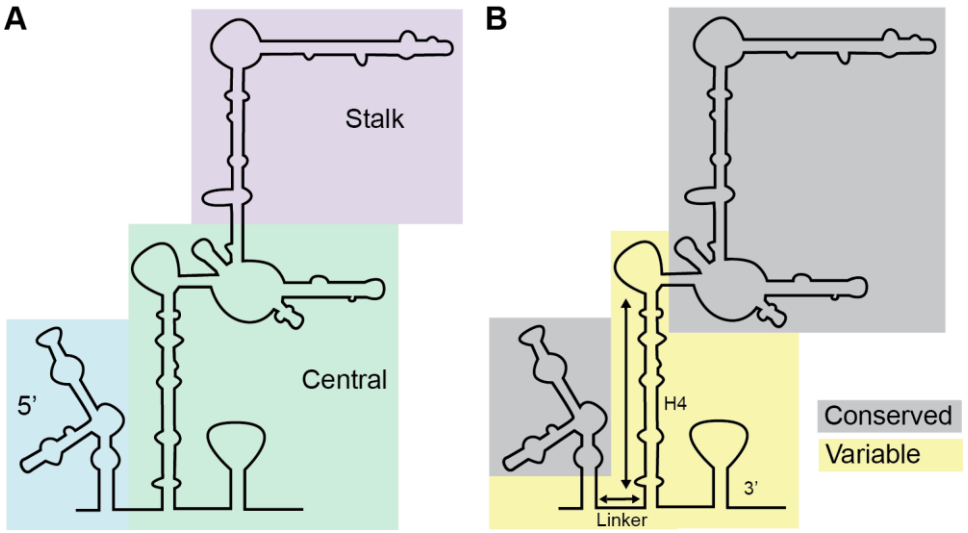


Figure 3.11: Modular domains and variable elements in the *COOLAIR* distal secondary structure. (A) The distal *COOLAIR* secondary structure can be divided into three main modular domains: 5' (blue), central (green), and stalk (purple). (B) The distal *COOLAIR* secondary structure can be divided into highly conserved (grey) or more variable (yellow) regions. Variation has been observed in the length of H4 and the linker region (arrows), and in how structured the 3' region is.

The role of the distal *COOLAIR* transcript in sense regulation is less well understood. High sense expression correlates with higher *COOLAIR* distal read-through (relative to proximal) as the *FLC* locus is in an active chromatin state; this suggests that the proximal is the stronger negative regulator. Yet changes in distal splicing drive *FLC* expression differences and later flowering time in natural populations (Li *et al.*, 2015). Strong conservation of distal structural elements, from *A. thaliana* to *B. rapa*, support that they are involved in flowering time regulation and provide insight into the distal regulatory mechanism. The 5' and stalk domains are highly conserved across all accessions and species, indicating that these may be critical for normal *COOLAIR* function (Fig. 3.11B).

The conserved H8/H9 structure, flanked by a robust and complicated secondary structure unit that shows covariance (first r-turn plus multi-way junction) suggests an important functional role, reinforced by the finding that this region associates with *FLC* chromatin in chromatin isolation by RNA purification (ChIRP) experiments (Csorba *et al.*, 2014). The 3' end and H4 (within the central domain) and the linker (between 5' and central domains) are less conserved and therefore more variable (Fig. 3.11B). This could indicate lack of function in these regions. Alternatively, it may be that variation here has been positively selected for under different environmental conditions over evolutionary time (Pang *et al.*, 2006). Indeed, the crystal structure of the eukaryotic ribosome revealed a highly-conserved core, along with separate more variable regions that allow for adaptation (Ben-Shem *et al.*, 2010).

This idea is particularly interesting with respect to the length and stability of helix H4. From functional (Li *et al.*, 2015) and structural (Hawkes *et al*., 2016) analysis, H4 and the associated first r-turn could be important in the regulation of *FLC* transcription in the warm. Its length and stability are significantly altered by the SNP responsible for the Var2-6 late flowering phenotype. The helix is also shorter and/or less stable in *A. alpina*, *E. salsugineum* and *B. rapa*; these might therefore behave more similarly to the Var2-6 than the Col-0 *COOLAIR*. It may be that length/presence of H4 is directly implicated in *COOLAIR* function, or that it alters accessibility to other domains involved in binding. For example, we speculate that the two rare r-turn motifs could play a role in ligand- or protein-binding; an r-turn in the U6 small nuclear ribonucleoprotein acts as a receptor for a protein (Montemayor *et al.*, 2014), whereas in the Pistol ribozyme the r-turn is a receptor for a pseudo-knot interaction (Ren *et al*., 2016). Identification of *COOLAIR* interacting protein complexes will help to determine whether this is correct. Despite sequence polymorphisms and variation in length, H4 is maintained in all species. Variation in H4 may be positively selected for under complementary environmental conditions, whereas complete disruption may be selected against if critical for any regulation by *COOLAIR* (whether weak or strong), without which the plant loses an extra level of regulatory control. Whilst a role for distal *COOLAIR* structure in the cold-induced epigenetic silencing of *FLC* is perhaps less likely because the two perennial species (*A. alpina* and *A. lyrata*) do not exhibit distinct structural features, the Var2-6 data are supportive of a role in setting initial levels of *FLC* expression in the warm. Further *COOLAIR* studies, including deletion of the first r-turn and compensatory mutations, will help to solve these structure-function relationships.

In summary, the global secondary structure of the distal *COOLAIR* has withstood evolutionary selection, despite low sequence similarity, while variation in other regions – such as H4 – has potentially allowed adaptation to a changing environment. These evolutionary analyses have helped us to move a step closer to understanding the role of *COOLAIR* in establishing expression levels of the *FLC* floral repressor. This further provides an important paradigm for genomic studies of lncRNAs, demonstrating how transcription and structure can be maintained despite low sequence conservation. We provide an example of a long non-coding RNA that is not transcriptional noise, rather it appears to be a useful tool in adaptation to different environments.

4 *FLC* and *COOLAIR* dynamics in *Brassica*

4.1 Introduction

After divergence from *Arabidopsis* approximately 43 million years ago (Ma), the ancestral *Brassica* underwent genome-wide chromosome triplication and subsequent duplication events to give rise to four copies of *FLC* (Rana *et al.*, 2004; Okazaki *et al.*, 2007; Beilstein *et al.*, 2010; Zou *et al.*, 2012). These four ancestral copies diversified over time, representing the four major *FLC* clades identified in *Brassica* species: *FLC1*, *FLC2*, *FLC3* and *FLC5* (Schranz *et al.*, 2002; Okazaki *et al.*, 2007). They have since diversified within species, i.e. after the *Brassica rapa*/*oleracea* split (Cheng *et al*., 2014). Tandem duplications of *FLC1* on chromosome 9 in *B. oleracea* – *BoFLCC9a*, *BoFLCC9b* and *BoFLCC9c* – have recently been reported (Chalhoub *et al.*, 2014; Rachel Wells, personal communication). The allotetraploid *Brassica napus* therefore contains four *FLC* copies from its *B. rapa* parent (the A genome) and six from its *B. oleracea* parent (C genome) (Tadege et al, 2001; Udall *et al.*, 2005).

The ancestral *Brassica* was presumably hexaploid, with six copies of each chromosome, in contrast to the diploid ancestral *Arabidopsis* (Lysak *et al.*, 2005). Polyploidy can confer selective advantages, such as increased organ size (polyploid plants are generally bigger) or improved adaptations to environmental pressures (Pires *et al.*, 2004). Multiple copies mean that there is a higher likelihood of mutation at any one *FLC* locus, a lower chance that a mutation will be lethal, due to redundancy between copies, and so a higher likelihood that mutations will be retained (Osborn *et al.*, 2003; Bomblies and Madlung, 2014). Immediate advantages can also be conferred to the new polyploid via alterations in gene regulation and dosage (epigenetic changes) and chromosomal re-arrangements (Osborn *et al.*, 2003 Gaeta *et al*., 2007; Cheng *et al*., 2013). *B. rapa* and *B. oleracea* are considered paleopolyploids, because although they are now diploid they evolved from a hexaploid ancestor. Reversion from polyploid to diploid can occur over deep time as a result of gene diversification or loss and chromosome reduction (Lysak *et al.*, 2006). Indeed, 35% of genes thought to be present when genome triplication occurred have now been lost (Town *et al.*, 2006). Polyploidy could have had a significant impact not only on divergence of FLC proteins but on *COOLAIR* long non-coding (lnc)RNAs. lncRNAs are less constrained than protein-coding sequence and therefore can evolve rapidly (Pang *et al.*, 2006; Kutter *et al.*, 2012; Nitsche *et al.*, 2015). In *Brassica* species, *COOLAIR* evolution may have been further accelerated by ancient polyploidy.

Similar gene architecture and high protein-coding sequence identity between *B. rapa* and *B. oleracea FLC* orthologues with *A. thaliana FLC*, 81.8-84.6% and 83-88% respectively, support evolutionary conservation of the FLC protein in *Brassica* (Schranz *et al.*, 2002; Razi *et al.*, 2008). Similarly, *B. napus FLC* copies have the same exon-intron architecture as their *B. rapa*/*B. oleracea* orthologues, and 95-100% sequence similarity with these in coding regions (Zou *et al.*, 2012). High *FLC* expression levels are associated with late flowering biennial cultivars and lower expression levels with annual early flowering cultivars (Kole *et al.*, 2001; Kim *et al.*, 2007). Overexpression of *B. rapa FLC* constructs delayed flowering in Chinese Cabbage and in *A. thaliana FLC*-null lines (Kim *et al.*, 2007), and five independent *B. napus FLC* transgenes delayed flowering in *A. thaliana* (Tadege *et al.*, 2001). These experiments suggest that *FLC* controls flowering time in a dosage-dependent manner in *Brassica*, as in *Arabidopsis*. *Brassica FLC* vernalization response may also work via a similar mechanism. Several groups have recorded a decrease in transgenic and endogenous *FLC* transcript levels during vernalization in annual and biennial *Brassica* cultivars (Kole *et al.*, 2001; Tadege *et al.*, 2001; Lin *et al.*, 2005; Kim *et al.*, 2007; Zhao *et al.*, 2010; Zou *et al.*, 2012). Downregulation of *FLC* can be quantitatively correlated with cold exposure and timing of flowering (Kim *et al.*, 2007). This correlates with an increase in *FT* transcript level in the shoot apex, although not in the leaves (Lin *et al.*, 2005). This suggests a similar mechanism whereby *FLC* represses floral integrator genes in non-vernalized conditions.

COOLAIR antisense transcripts have been conserved at *FLC* orthologues in other members of the Brassicaceae family, including *B. rapa* (Castaings *et al.*, 2014; Hawkes *et al.*, 2016; Li *et al.*, 2016). In Chapter 3, *COOLAIR* was detected from one of four *FLC* copies in *B. rapa* (*BrFLCA3a*). It shares similar architecture with *COOLAIR* from *A. thaliana*, with proximally and distally polyadenylated transcripts, conserved splice sites and conserved secondary structure. Upregulation of *COOLAIR* in the cold has been reported at *BrFLCA2* (Li *et al.*, 2016). It is therefore possible that *COOLAIR* regulates *FLC* sense expression before and during cold in *Brassica* as it does in *Arabidopsis*. Intra-plant allelic variation between *FLC* copies should be a strong enhancer of flowering time diversity (Schranz *et al.*, 2002). Multiple functional copies of *FLC* may have been retained through evolution to add an extra level of quantitative control to *Brassica* flowering time, with each contributing to the decision to switch to flowering (Schranz *et al.*, 2002). It is not clear, however, whether the contribution from each *FLC* copy is equal. There may instead be one (or a few) major player(s) in flowering time control, with allelic differences at that copy overriding the effects of other copies. For example, major vernalization-responsive QTLs

in *B. rapa* and *B. napus* have been mapped to chromosome 10, which contains the *FLCA10* copy (Osborn *et al.*, 1996); whereas, late-early flowering time QTLs across diverse *B. rapa* cultivars have been mapped to *FLCA2* (Lou *et al.*, 2007; Zhao et al, 2010). Similarly, flowering time QTL for late-early flowering responses in *B. oleracea* and *B. napus* have been mapped to *FLC2* (Okazaki *et al.*, 2007; Zou *et al.*, 2012; Irwin *et al*., 2016). In *Arabidopsis*, *cis* and *trans* mutations can work through *COOLAIR* or through regulators of *COOLAIR* to alter sense expression (Liu *et al*., 2010; Marquardt *et al*., 2014; Li *et al.*, 2015). With the potential for even greater natural variation at *FLC* within and between *Brassica* cultivars, we propose that some expression differences will be driven by *COOLAIR*.

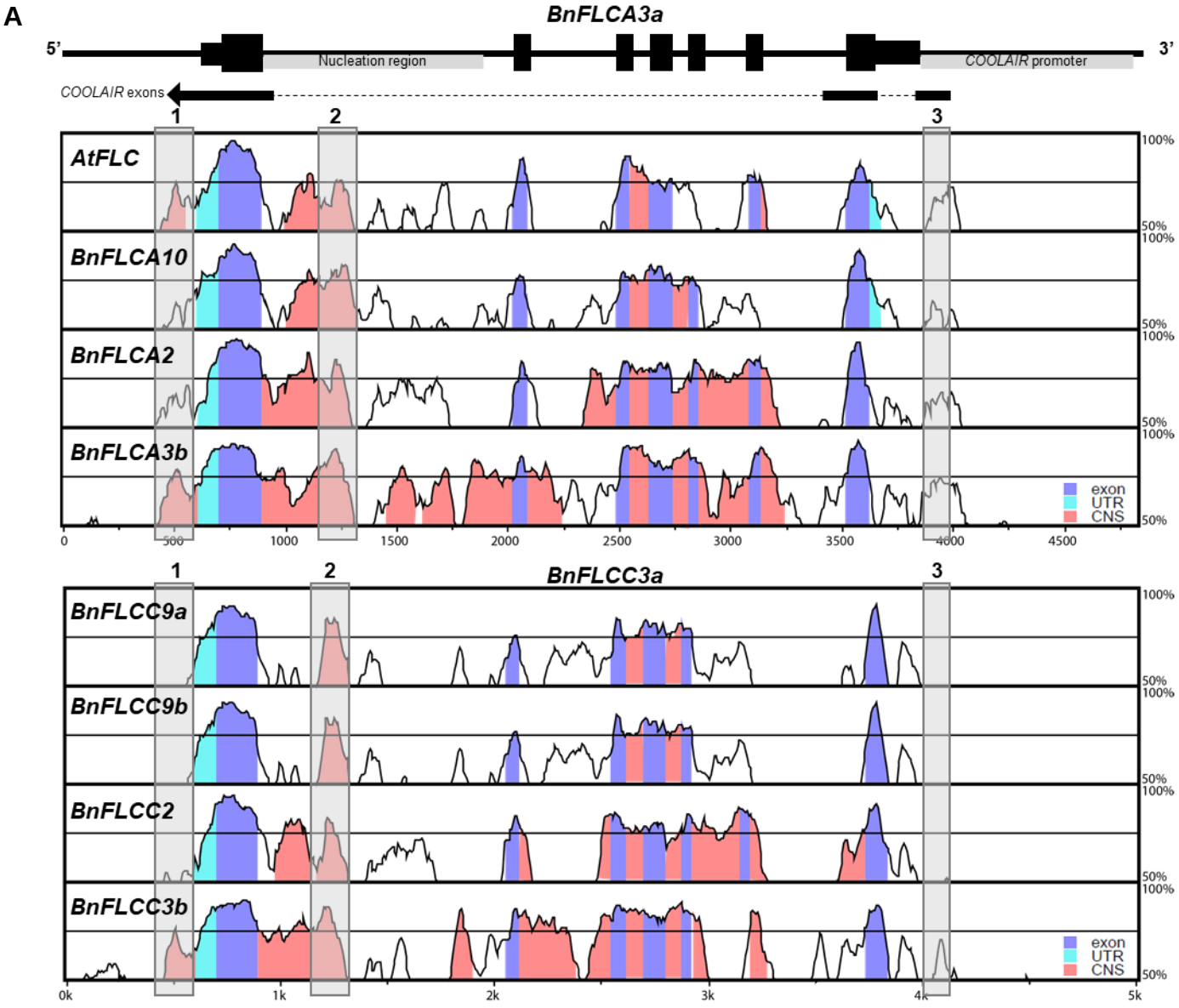
Only 3% of *Arabidopsis* flowering time genes have not to date been found to have a *Brassica* paralogue, with 30% having more than three paralogues in *Brassica rapa* (Xiao *et al.*, 2013). This suggests strong conservation of flowering time genes between *Brassica* and *Arabidopsis*, with multiple copies preferentially retained (compared with non-flowering time genes) (Schiessl *et al*., 2014). Potentially similar pathways of flowering time control are complicated by multiple copies of *FLC* and multiple copies of the proteins it interacts with (Schiessl *et al*., 2014). We do not know which copy interacts with which regulator, or whether they have equal capacity to interact with them all. *COOLAIR* transcripts are transcribed from at least two *FLC* loci in *Brassica rapa* (Li *et al.*, 2016; Hawkes *et al.*, 2016). If the regulatory function of *COOLAIR* has been conserved from *Arabidopsis* to *Brassica*, this begs the question of whether *COOLAIR* acts in *cis* to regulate the *FLC* at the locus from which it was transcribed, in *trans* to regulate other *FLC* loci, or both. In the same way there may be a master *FLC* regulator of flowering time, there could be a master *COOLAIR* regulator of all *FLC* copies, with other *COOLAIR* transcripts becoming redundant over evolutionary time. Li *et al.* (2016) were only able to detect antisense transcripts from *BrFLCA2*; consequently, they proposed that these were regulating the other homologues in *trans*. We now know *COOLAIR* is transcribed from at least one other copy in *B. rapa* (*BrFLCA3a*). Here, we set out to better understand how sense and antisense transcripts at *FLC* fine-tune flowering time in *Brassica* species.

4.2 Results

4.2.1 Brassica FLC *homologues have distinct nucleotide sequences*

The *Brassica* A and C genome *FLC* homologues share their canonical gene structure with *A. thaliana FLC* (*AtFLC*): seven exons and six introns, with the largest intron (intron 1) between exons 1 and 2 (Schranz *et al.*, 2002; Razi *et al.*, 2008; Zou *et al.*, 2012). The length of intron 1 varies across homologues and is smallest in *FLC3* (*FLCA3a* and *FLCC3a*) and largest in *FLC1* (*FLCA10* and *FLCC9*); there is also variation in the length of intron 6 (Schranz *et al.*, 2002; Razi *et al.*, 2008). Here, we analyse *FLC* homologues from *B. napus* Cabriolet. The A genome *FLC* homologues were aligned with *BnFLCA3a* and the C genome homologues with *BnFLCC3a* (Fig. 4.1A) to look for conserved regions. *FLC3* homologues were selected to avoid comparison of insertion-deletions in intron 1 (*FLC3* contains the shortest first intron). Note that intron 1 size therefore cannot be ascertained from Fig. 4.1A.

There is a conserved ~200 bp block towards the 5' end of intron 1 (grey box 2, Fig. 4.1A). This is within the nucleation region and contains the RY motifs, required for initiation and maintenance of *FLC* shut-down during vernalization in *Arabidopsis* (Sheldon *et al.*, 2002; De Lucia *et al.*, 2008; Qüesta *et al.*, 2016). *Brassica FLC* homologues may therefore undergo a similar mechanism of cold-response. Sequence similarity is highest in the protein-coding sense exons, indicating conservation of FLC protein structure and function. Exon conservation is, however, relatively low in some homologues i.e. there is low conservation of exon 6 between *BnFLCA10* and *BnFLCA3a*. Exon 1, which contains the functional MADS box domain, is the most highly conserved (Michaels and Amasino, 1999). The 5'UTR, adjacent to exon 1 and containing part of the *COOLAIR* distal terminal exon, is also represented by a conserved block (coloured turquoise), although sequence similarity drops 5' of this (grey box 1, Fig. 4.1A). *BrFLCA3a* did not share the ~100 bp conserved block at the *COOLAIR* promoter region with *Arabidopsis* in the analysis in Chapter 3 (Fig. 3.2; grey box 3 in Fig. 4.1A); likewise, there is low homology in this region between *BnFLC3* and the other *FLC* homologues. Perhaps, *FLC3* is the exception and this region has been conserved in other homologues. When we compare *BnFLCA10* with the other A and C genome *BnFLC* homologues using the same method, we still observe low homology in this region (Fig S2). Direct comparison between each homologue and *AtFLC* suggests some conservation in this region with *BnFLCA2*, *BnFLCA3a* and *BnFLCA3b* but not with the other homologues (Fig S3).



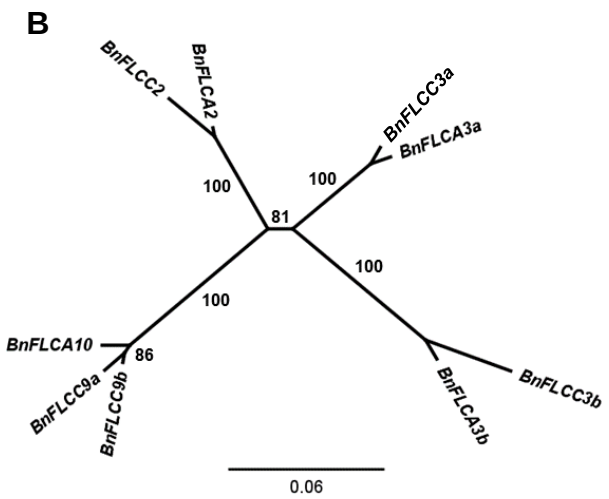


Figure 4.1: Sequence conservation of *Brassica FLC* homologues. (A) Sequence conservation of *A. thaliana* and *B. napus* (Cabriolet) A genome *FLC* homologues compared with *BnFLCA3a* (top), and of *B. napus* Cabriolet C genome *FLC* homologues with *BnFLCC3a* (bottom). MLAGAN progressive pair-wise alignments with translated anchoring performed and visualized with the mVISTA online tool, with a window size of 100 bp and cut-off criteria of 70% sequence identity (Brudno *et al*. 2003; Frazer *et al*., 2004). Coloured areas indicate stretches of homology greater than 75% identity at the nucleotide level. *FLC* sequences are from 0.7 kb upstream of ATG start to 1.2 kb downstream of translational stop, apart from *FLCC3b* which only extended ~0.7 kb downstream of translational stop. Numbered grey boxes highlight areas of interest. CNS = conserved non-coding sequence. (B) Phylogenetic tree created from MUSCLE multiple sequence alignment of *B. napus* Cabriolet *FLC* mRNA using the HKY neighbour-joining model (Geneious R7). Topology tested with 1000 bootstrap replicates (consensus tree shown with % consensus support labelled).

Differences between these results probably reflects variation in pair-wise versus multiple sequence alignment using the mLAGAN tool. Castaings *et al.* (2014) suggested this block had been evolutionarily conserved to maintain *COOLAIR* transcription. Sequence divergence could therefore have led to *COOLAIR* loss in some of the *Brassica FLC* homologues, or changes in *COOLAIR* function.

There are four distinct *Brassica FLC* clades representing the early genome triplication event that gave rise to *FLC1*, *FLC2* and *FLC3*, and the subsequent duplication of *FLC3* that gave rise to *FLC5* (Tang *et al*., 2012). Phylogenetic analysis of nine *FLC* homologues in *B. napus* support this relationship, with four main branches representing the four clades, divided at 81% certainty between *FLC3* and *FLC5* versus *FLC1* and *FLC2* (Fig. 4.1B). This division may further represent the 'two-step fractionation' hypothesis of genome triplication, with *FLC1* and *FLC2* belonging to subgenomes I and II, which are thought to have preceded subgenome III (Tang *et al*., 2012). The A and C genome homeologues split at the end of each main branch, representing their much more recent divergence within *B. rapa* and *B. oleracea*. *FLCC9a* and *FLCC9b* arose from an independent duplication of *FLCC9* in *B. oleracea*, and are therefore positioned more closely to each other than to *FLCA10* (Chalhoub *et al*., 2014). There is thought to be an additional *BnFLCC9c* copy, not present in this data set (Rachel Wells, personal communication). Each *FLC* homologue is distinct in terms of primary nucleotide sequence, with each clade as different from each other as each is from *AtFLC* (Table 4.1). Full *FLC* nucleotide sequence similarity between clades ranges between 27.9-46.9%, which is lower than that found between species in Chapter 3. Full *FLC* nucleotide sequence similarity between homeologues is higher, but still low at between 59.3-77.3%. These results, together with Fig. 4.1, suggest that vast tracts of the *FLC* gene are under no selective pressure.

% identity	*FLCA10*	*FLCC9a*	*FLCC9b*	*FLCA2*	*FLCC2*	*FLCA3a*	*FLCC3a*	*FLCA3b*	*FLCC3b*
FLCA10		69.0	63.3	33.7	33.4	39.6	38.3	43.0	33.1
FLCC9a	69.0		63.5	33.6	32.3	39.9	38.0	40.8	31.5
FLCC9b	63.3	63.5		29.4	27.8	34.4	33.3	36.6	29.4
FLCA2	33.7	33.6	29.4		70.4	46.9	45.3	39.4	28.5
FLCC2	33.4	32.2	27.8	70.4		44.3	43.8	38.7	27.9
FLCA3a	39.6	39.9	34.4	46.9	44.3		77.3	42.1	30.2
FLCC3a	38.3	39.0	33.3	45.3	43.8	77.3		42.0	30.7
FLCA3b	43.0	40.8	36.6	39.4	38.7	42.1	42.0		59.3
FLCC3b	33.1	31.5	29.4	28.5	27.9	30.2	30.7	59.3	

Table 4.1: % nucleotide identity of genomic regions from nine *B. napus* Cabriolet *FLC* homologues. % nucleotide identity based on Muscle multiple sequence alignment (Geneious R7). *FLC* sequences were from 0.7 kb upstream of ATG start to 1.2 kb downstream of translational stop, apart from *FLCC3b* which only extended ~0.7 kb downstream of translational stop.

4.2.2 Brassica FLC *homologues have distinct expression patterns*

Given high sequence variation, it is not surprising that *Brassica FLC* homologues exhibit unique expression patterns. *FLC* sense (spliced mRNA) expression was determined under non-vernalized conditions for eight/ten homologues in twelve (six winter, six spring) *B. napus* cultivars (Fig. 4.2A and B). *BnFLCC9a* was included in this study but not *BnFLCC9b* or *BnFLCC9c*; the latter two proved difficult to design specific primers for due to their high sequence similarity with each other and with *BnFLCC9a*. Non-vernalized expression represents the 'normal' or 'starting' levels of *FLC* in a plant. This helps to determine vernalization requirement and flowering time with and without vernalization. Six winter (blue) and six spring (orange/pink) cultivars were analysed in two separate experiments, Fig. 4.2A and Fig. 4.2B. Absolute values vary by 10-fold between the experiments due to use of redesigned more efficient (97% versus 92%) GAPDH* reference primers for expression analysis in Fig. 4.2B; the original GAPDH primer set is still within the recommended 90-110% range, however (Life Technologies, 2012). The experiments were independent and so differences may also reflect plant growth conditions. Consequently, expression levels are not directly comparable between Fig.4.2A and B, although differences within or between cultivars within each experiment are. These experiments revealed that *FLC* sense expression varies within and between cultivars.

Total *FLC* expression (across all homologues) was generally higher in winter than in spring lines. This supports the idea that each *FLC* copy individually represses flowering, with absolute flowering time the sum contribution from each. If true, we could expect flowering time across these 12 cultivars to reflect the sum of their *FLC* copies, with W150 flowering later than W42 for example. This assumes all other factors are constant (i.e. the same *trans*-regulators are present) and that the potency of each protein is the same, which is unlikely to be the case. In fact, none of the four winter lines in Fig. 4.2A had flowered by 120 days under non-vernalized conditions (Fig. 4.2C; flowering time data produced by Eleri Tudor and used with permission). This could suggest that the quantitative sum of *FLC* expression was high enough to prevent flowering and that these lines have an obligate requirement for cold; alternatively, differences may have become apparent beyond 120 days. Two out of three plants of the Major winter cultivar flowered late but within 120 days; this is interesting because Major had ~2-fold lower *BnFLCA10* than *BnFLCA3a* expression, whereas in other winter lines *BnFLCA10* expression was equal to or higher than *BnFLCA3a*. Spring lines flowered rapidly and within a few days of each other (Fig. 4.2C). S48 flowered marginally later than S69 or S135; it had similarly low overall *FLC* expression levels, but *BnFLCC3a* expression was higher in this line than in the others.

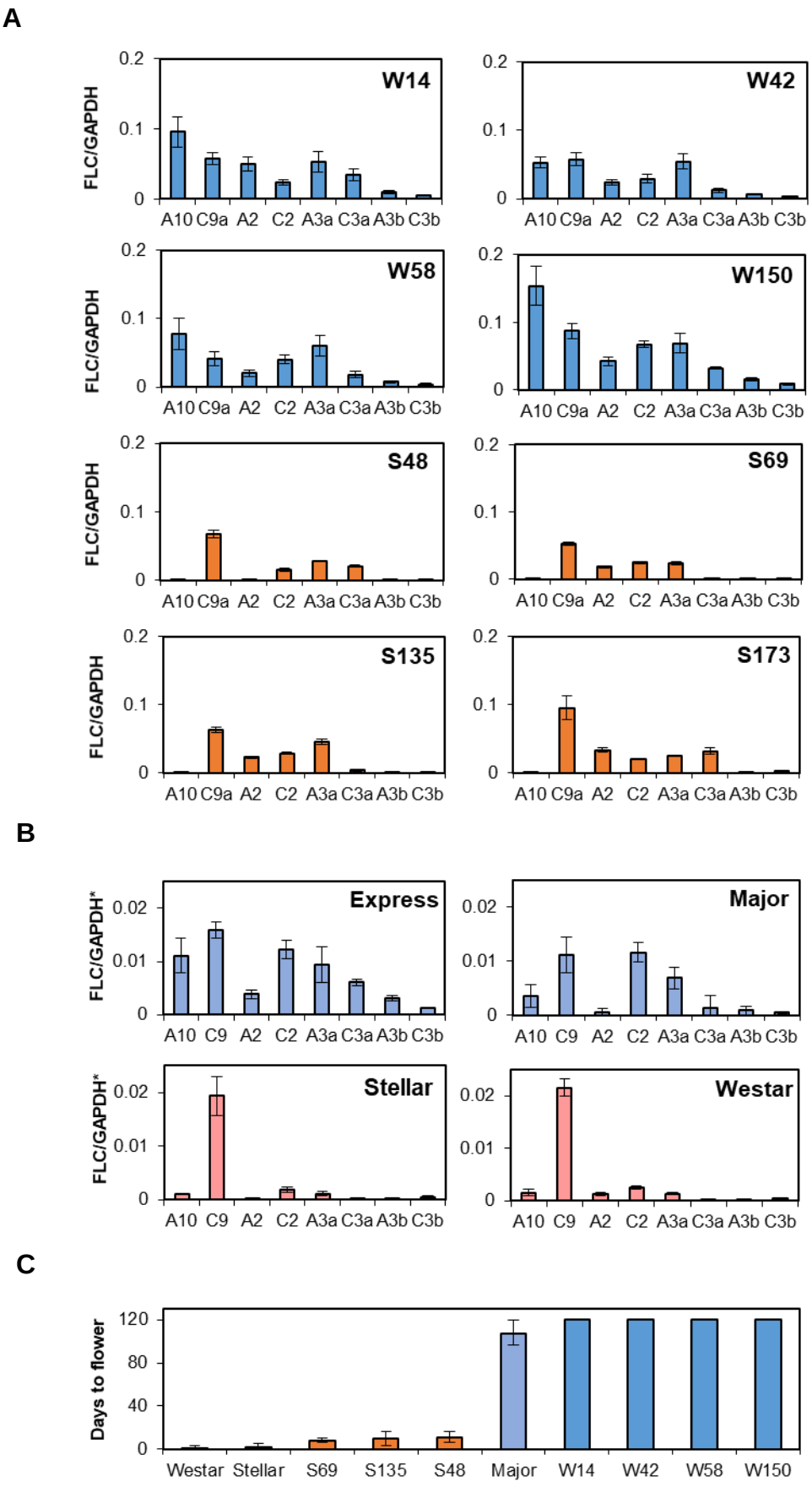


Figure 4.2: *FLC* expression and flowering times of non-vernalized *B. napus* cultivars. (A and B) Expression of spliced sense *FLC* mRNA in six winter (blue) and six spring (orange/pink) cultivars. (A) and (B) were separate experiments, with (A) normalised to *GAPDH* primers and (B) normalised to *GAPDH** primers (Table 2.12). Error bars are 1 St Dev. from the mean for three biological replicates. (C) Flowering times of ten *B. napus* cultivars (days to flower). Error bars represent range for three plants per cultivar. Flowering time trial and data collection carried out by Eleri Tudor in 2014 and used with permission. W14, W42, W58 and W150 had not flowered by 120 days.

We predict that some copies are stronger regulators of flowering time than others. *BnFLCA10*, *BnFLCC9a*, *BnFLCA3a* and *BnFLCC2* were expressed at the highest levels across the six winter lines. Expression of three of these four, *BnFLCA10*, *BnFLCC2* and *BnFLCA3a*, was lower in the six summer lines. Variation in these three homologues may therefore be responsible for flowering time differences in spring versus winter lines, i.e. in determining whether or not the plant has a requirement for cold. *BnFLCC9a* expression, although high in winter lines, was also high in the six early flowering spring lines. In fact, *BnFLCC9a* expression was often higher than all other copies in the spring lines: approximately 10-fold higher in Stellar and Westar and 2-fold higher in S48, S69 and S173. This suggests that the BnFLCC9a protein (assuming it is also highly expressed) is no longer involved in floral repression. Expression of *FLC5* (*BnFLCA3b* and *BnFLCC3b*) was low in both winter and summer lines. If expression levels are directly proportional to the amount of FLC protein produced, we would expect these to be less important for floral repression. *BnFLCA2* and *BnFLCC3a* expression levels were higher than *FLC5*, but whereas *BnFLCC3a* expression was lower in spring than in winter lines, *BnFLCA2* expression was similar in both. The candidates that seem most interesting in terms of setting initial *FLC* levels in the warm and giving rise to a late or early flowering phenotype are therefore *BnFLCA10*, *BnFLCC2*, *BnFLCA3a* and *BnFLCC3a*. Expression differences between homeologues [with their high sequence similarity] implies that a low number of polymorphisms can contribute to changes in expression. This is supported by work in *Arabidopsis thaliana*, where single *cis* non-coding polymorphisms can alter *FLC* sense expression and flowering time (Coustham *et al.*, 2012; Li *et al.*, 2015; Qüesta *et al.*, 2016). Some differences in expression are likely due to changes in *trans*-regulators. As there are multiple copies of *FLC* regulators, it may be that each works with a specific copy (or copies) of *FLC*.

A vernalization time-series of two winter and two spring cultivars provided further insight into regulation of flowering time by *FLC* in *Brassica napus* (Fig. 4.3). Three plants for each cultivar were treated to two, four and six weeks cold, flowering times recorded and *FLC* expression analysed from leaf tissue samples before, during and after cold treatment (Fig. 4.3A and B). Flowering times were also compared with R018, a *B. rapa* spring variety that has no cold requirement, and DH1012, a *B. oleracea* semi-winter variety with a small cold requirement. Whereas the two winter lines did not flower within 80 days without cold treatment, the two spring lines flowered after ~10 days. Subsequent cold treatment did not speed up flowering in the *B. napus* or *B. rapa* spring lines, indicating that they do not have a cold requirement. This correlates with the low levels of all *FLC*

homologues observed in these lines (with the exception of *BnFLCC9a*), reported above. The two winter lines did not flower within 80 days with two weeks cold, but did flower after four and six weeks cold. Flowering was still much later than for the spring lines, at 30-50 days after cold treatment, indicating that vernalization requirement was not fully saturated. There were minor differences in flowering times between the two winter cultivars, Express and Major, despite Major having relatively lower non-vernalized levels of all *FLC* homologues except *BnFLCC2* (Fig 4.2) – this may be because non-vernalized *FLC* levels are important for cold requirement, but less important for cold response.

After two weeks vernalization, *BnFLCA10* and *BnFLCA3a* expression levels are downregulated in winter lines; they remain low at 15 and 30 days after cold (Fig. 4.3C). Some reactivation of *BnFLCA10* and *BnFLCA3a* may occur in Major 15 days after cold treatment, but levels stabilise by 30 days after cold rather than reactivating further. A similar pattern of reactivation and stabilisation is observed for *BnFLCA2* and *BnFLCA3b* sense transcripts in the winter lines, where starting *FLC* expression levels were low. Expression of the A genome *FLC* transcripts were already low in spring lines and remain at a similar level during and after cold treatment; this suggests they cannot be further downregulated by cold. These data for the A genome follow typical *AtFLC* downregulation patterns in the cold, although it is surprising that all copies are low or stably repressed after only two weeks vernalization and without flowering of the winter lines.

The C genome responds differently. *BnFLCC2* and *BnFLCC3a*, although initially downregulated during cold, reactivate by 15 days and return to non-vernalized levels or higher by 30 days after cold. This suggests that it may be *BnFLCC2* and/or *BnFLCC3a* preventing the winter lines from flowering after two weeks cold, as we know the A genome *FLC* copies have been deactivated. *BnFLCC3b* is expressed at a low level before and after cold in both spring and winter lines and therefore does not look to be playing a role in floral repression. *BnFLCC9a* is also downregulated during cold treatment and reactivates after cold. Interestingly, this reactivation is even more extreme in the two summer lines, where *BnFLCC9a* was highly expressed before vernalization and where the plants are flowering by the 15 days time-point. This supports the hypothesis that *BnFLCC9* is not a floral repressor; high activity during flowering could instead signal a role in floral development or seed set. Similar expression patterns are observed after four weeks of cold treatment, with stable downregulation or low levels of A genome *FLC* copies and reactivation of C genome *FLC* copies (with the exception of *BnFLCC3b* which remains at a low level throughout) (Fig. 4.3D).

A

Cultivars	Vernalization	Sampling times	Plants
R018	2, 4, 6 weeks	T0, T15, T30	3
DH1012	2, 4, 6 weeks	T0, T15, T30	3
Express	2, 4, 6 weeks	T0, T15, T30	3
Major	2, 4, 6 weeks	T0, T15, T30	3
Stellar	2, 4, 6 weeks	T0, T15, T30	3
Westar	2, 4, 6 weeks	T0, T15, T30	3

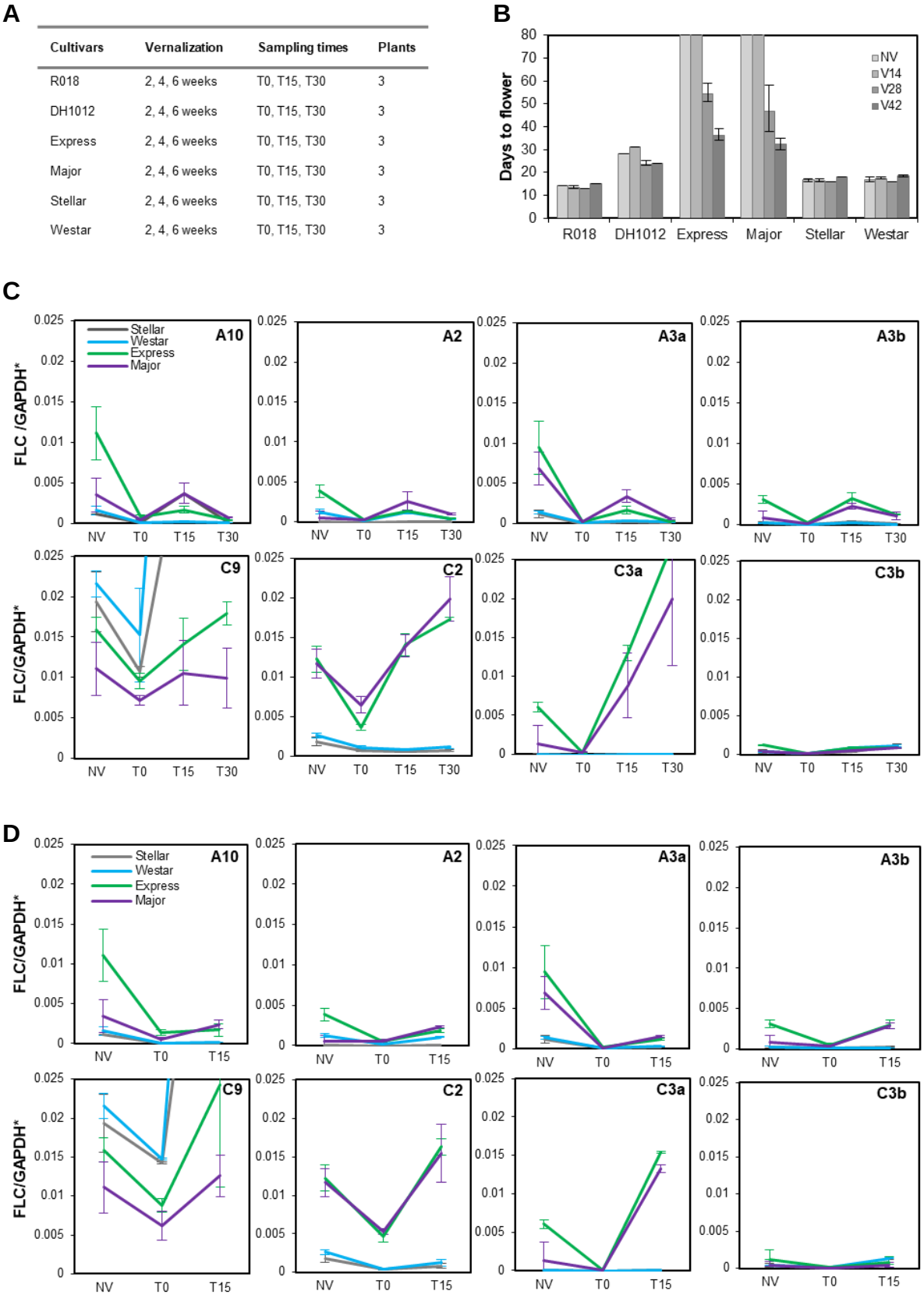


Figure 4.3: Vernalization time-series for four *B. napus* cultivars. (A) Experimental details. (B) Flowering times for *B. rapa* R018, *B. oleracea* DH1012, and *B. napus* Express, Major, Stellar and Westar cultivars (days to flower) under non-vernalized conditions (NV) and after two (V14), four (V28) and six (V42) weeks vernalization. NV and two weeks vernalized Express and Major had not flowered by 80 days. Error bars represent the range in flowering times for three plants per cultivar. (C and D) Expression of spliced sense *FLC* mRNA at NV and after two (C) or four (D) weeks vernalization for the four *B. napus* cultivars. qPCR runs with four-week T30 and six-week samples failed, hence these data are missing. Samples were normalized to the *GAPDH* reference gene (the *GAPDH** primer set). Error bars are 1 St Dev. from the mean for 2-3 biological replicates. C9 represents *FLCC9a*.

The winter lines in this experiment did flower, although not as early as the spring lines, suggesting vernalization requirement is not fully saturated. It is not clear from these data (which are missing T30 values due to qPCR failure) whether stabilization in *BnFLCC2* or *BnFLCC3a* expression at T30 is responsible for this difference [in flowering times with two weeks vernalized plants], although there is still strong reactivation of both at T15.

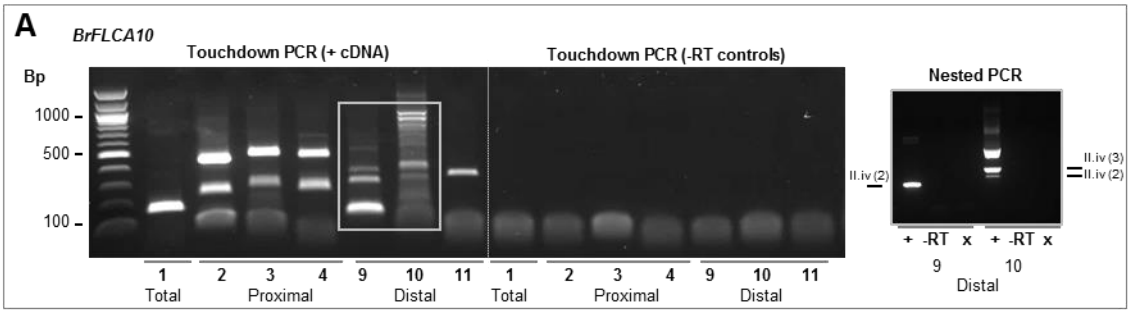
These data suggest that although the A genome may play a role in the initial setting of *FLC* levels in the warm (and whether there is a requirement for cold), it is the C genome that is important for determining vernalization response and the *amount* of cold required for flowering. This is interesting because it supports A genome predominance in the warm, with *BnFLCA10* as a critical determinant of spring or winter phenotype (Schiessl *et al*., 2017). In the cold, the A genome *FLC* copies are rapidly switched off, whereas the C genome copies not only reactivate, but appear to be expressed at higher levels than before cold. One possible explanation for their opposing patterns after cold is that there is cross-talk between the *FLC* homeologues. If the 'dominant' A genome homologues repress the C genome homologues before cold, then the downregulation of the A genome during cold would give rise to this exact phenomenon. In support of this idea, when we directly compare A and C genome homeologue expression patterns in winter lines, we note a mirror imagery between the two (Fig. S4). The exception is *FLC5*, where expression of both homeologues is low.

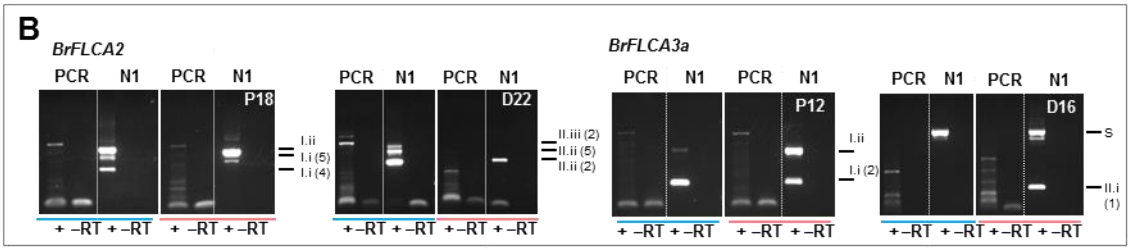
4.2.3 COOLAIR *is expressed from three of the four* Brassica FLC *clades*

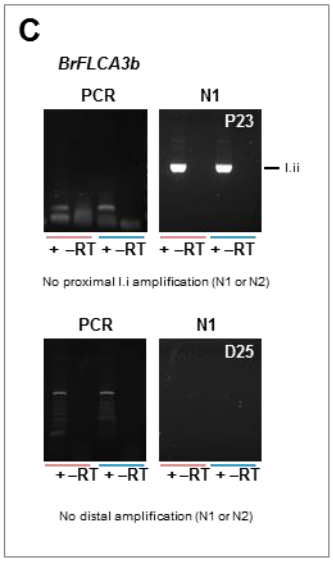
COOLAIR primers were designed for each of the *FLC* copies in *B. rapa* and *B. oleracea*, using *BrFLCA3a* and *AtFLC* antisense transcripts as a guide. Primers were designed to detect total spliceoforms, i.e. from the predicted first and last exons of the proximal and distal transcripts. A detection method of touchdown PCR followed by one or two nested PCRs was developed to ensure (1) specific amplification, (2) abundant product, (3) detection of low level transcripts, and (4) detection of all spliceoforms. The splice forms were broadly characterised based on their similarity to *A. thaliana* classes (Class I and Class II, Fig. 3.1), although new subdivisions have been added to account for the greater variation we see in *Brassica*, i.e. Class II.i (1-5) (see Fig. 4.7 for schematic of different *Brassica* splice transcripts). Results were semi-quantitative for the first (but not nested) PCR as cDNA samples were equalised for the *UBC* reference gene at amplicon level; this allowed interpretation of absence/prescence differences between vernalized and non-vernalized samples.

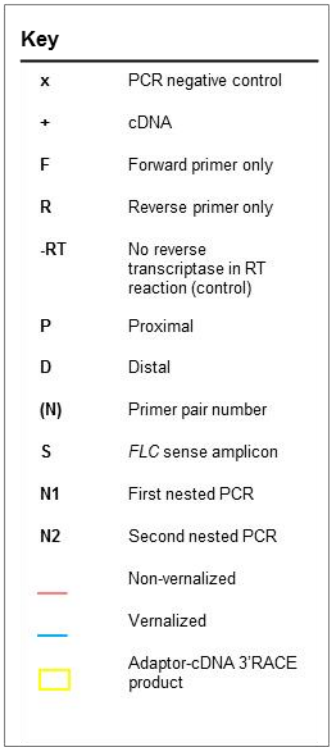
Proximal and distal *COOLAIR* transcripts were detected from *FLCA10*, *FLCA2* and *FLCA3a* from the *B. rapa* A genome (Fig. 4.4) and from their C genome homeologues in *B. oleracea* (Fig. 4.5). The unspliced proximal Class I.ii transcript, but no other proximal or distal transcripts, were detected from *FLC5* (*BrFLCA3b* and *BoFLCC3b*). We therefore propose that *COOLAIR* has been conserved in at least three of the four *Brassica FLC* clades. This is despite the lack of a conserved sequence block in the *COOLAIR* exon 1/promoter region and high sequence divergence between homologues. This supports the hypothesis presented in Chapter 3 that conservation of primary nucleotide sequence is not necessary for conservation of transcription, secondary structure or function. Cloning and sequencing of PCR products revealed multiple spliceoforms across and within homologues, with some detected only in the cold, some only in the warm, and some in both. These are discussed in detail in section 4.2.6.

The proposed existence of both 'proximal' and 'distal' transcripts in *Brassica* is based on spliceoform similarity to *Arabidopsis* classes. It is not, however, possible to tell from the above experiment where they terminate (and so if they are indeed proximal and distal transcripts). Polyadenylation sites were therefore mapped by 3'RACE from the *B. rapa FLCA10* proximal and distal terminal exons (Fig. 4.4D). *B. rapa* RNA (from non-vernalized leaf tissue) was ligated with an RNA-DNA hybrid adaptor molecule and used as a substrate for gene-specific RT-PCR, with adaptor-specific reverse primers (see methods). Total PCR products were cloned collectively, with up to 96 clones sequenced per reaction to sample spread in polyadenylation sites. Multiple 3'RACE products may indicate non-specificity of the reaction primers or the use of multiple polyadenylation sites. Additional forward and reverse primer-only controls were included to detect [and thus exclude] non-specific amplification from a single primer. Decrease in product size between Nest 1 and Nest 2 (amplitude dependent on the positions of the nested primers) also indicated a positive product.


A
BrFLCA10
Touchdown PCR (+ cDNA)
Touchdown PCR (-RT controls)
Nested PCR
Bp
1000
500
100
1 2 3 4 9 10 11 1 2 3 4 9 10 11
Total Proximal Distal Total Proximal Distal
II.iv (2)
II.iv (3)
II.iv (2)
+ -RT x + -RT x
9 10
Distal


B
BrFLCA2
BrFLCA3a
PCR N1 PCR N1
P18
D22
P12
D16
I.ii
I.i (5)
I.i (4)
II.iii (2)
II.ii (5)
II.ii (2)
I.ii
I.i (2)
S
II.i (1)
+ –RT + –RT


C
BrFLCA3b
PCR N1
P23
I.ii
+ –RT + –RT + –RT + –RT
No proximal I.i amplification (N1 or N2)
PCR N1
D25
+ –RT + –RT + –RT + –RT
No distal amplification (N1 or N2)


Key
x PCR negative control
+ cDNA
F Forward primer only
R Reverse primer only
-RT No reverse transcriptase in RT reaction (control)
P Proximal
D Distal
(N) Primer pair number
S FLC sense amplicon
N1 First nested PCR
N2 Second nested PCR
Non-vernalized
Vernalized
Adaptor-cDNA 3'RACE product

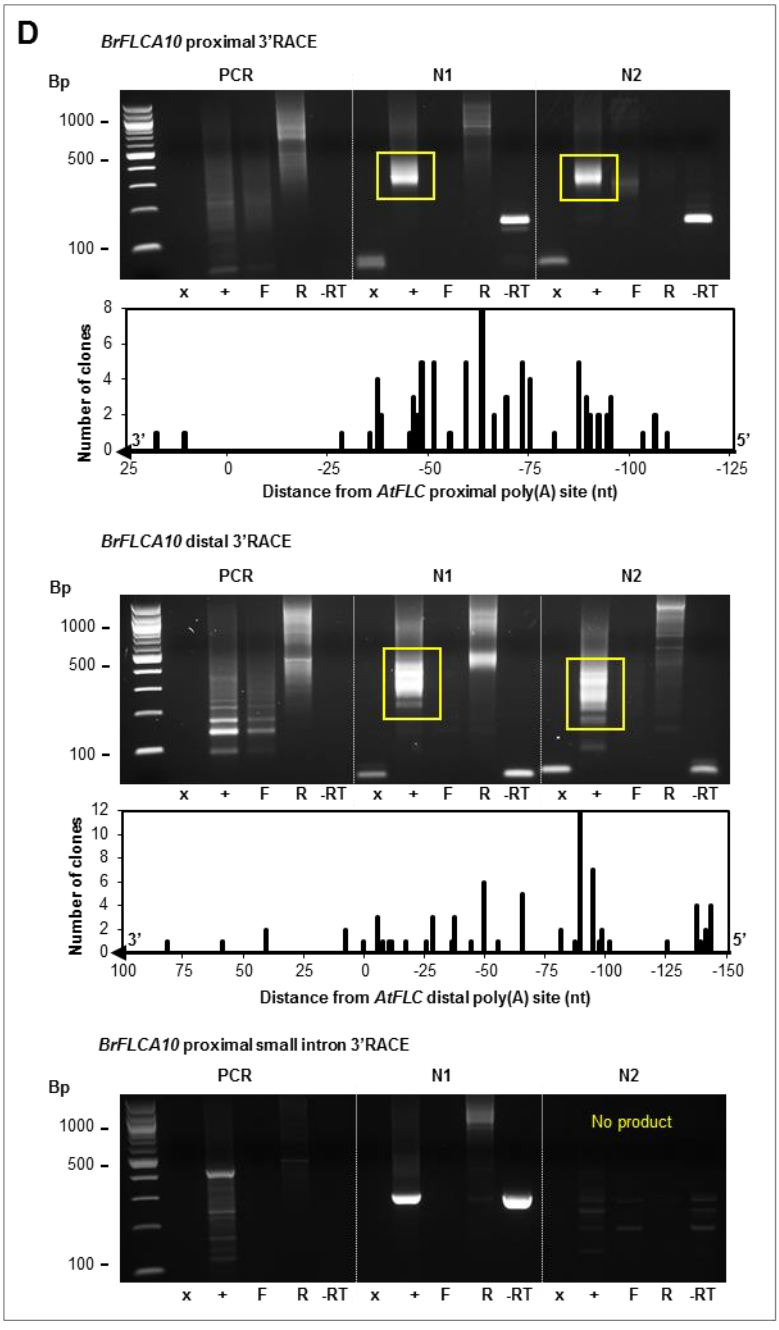

D
BrFLCA10 proximal 3'RACE
Bp
PCR N1 N2
1000
500
100
x + F R -RT x + F R -RT x + F R -RT
Number of clones
Distance from AtFLC proximal poly(A) site (nt)
3'
5'
BrFLCA10 distal 3'RACE
Distance from AtFLC distal poly(A) site (nt)
BrFLCA10 proximal small intron 3'RACE
No product

Figure 4.4: Characterisation of *COOLAIR* in *B. rapa FLC* homologues (previous page). (A, B, C) Touchdown and nested RT-PCR experiments probing for *COOLAIR* total, proximal and distal transcripts from *B. rapa* R018 *FLCA10* (A), *FLCA2* (B, left), *FLCA3a* (B, right) and *FLCA3b* (C) under non-vernalized (pink line) or two weeks vernalized (blue line) conditions. Initial touchdown PCR was normally followed by two nested PCRs (N1 and N2), where P(N) and D(N) refer to proximal and distal primer set numbers, respectively. Here, only initial PCR and N1 gel images are given, as (where used) N2 included no new information. Input cDNA was equalised for the *UBC* reference gene at amplicon level to give semi-quantitative results for the first PCR. Different splice variants have been categorised according to *A. thaliana* classes in Fig. 3.1, for full details see Fig. 4.7. (D) 3'RACE experiments probing for *B. rapa FLCA10 COOLAIR* proximal (top), distal (middle) and proximal unspliced (bottom) polyadenylation sites. Gel images show initial PCR of the cDNA (ligated to an adaptor molecule) followed by two rounds of nested PCRs (N1 and N2). Graphs detail position of polyadenylation sites relative to the canonical *A. thaliana* proximal or distal site (position 0), and the number of clones (out of up to 96 per experiment) mapped to that site.

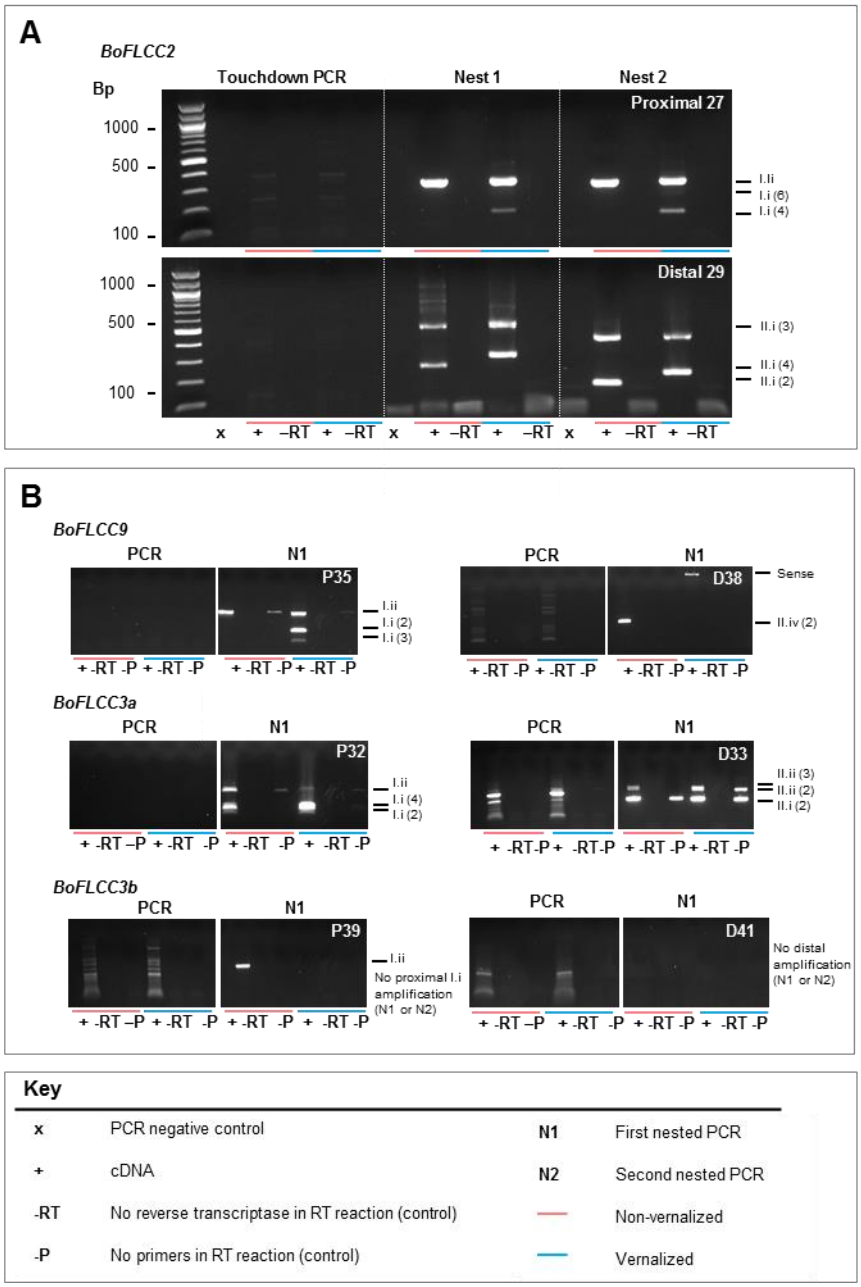


Figure 4.5: Characterisation of *COOLAIR* in *B. oleracea FLC* homologues. (A, B) Touchdown nested RT-PCR experiments probing for *COOLAIR* proximal and distal transcripts from *B. oleracea* PSB *FLCC2* (A), and *B. oleracea* DH1012 *FLCC9* (B, top), *FLCC3a* (B, middle) and *FLCC3b* (B, bottom) under non-vernalized (pink line) or two (B) to 10 (A) weeks vernalized (blue line) conditions. Initial touchdown PCR was followed by two nested PCRs (Nest 1 and Nest 2), where P(N) and D(N) refer to proximal and distal primer set numbers, respectively. In (B) only initial PCR and N1 gel images are given, as N2 included no new information. Input cDNA was equalised for the *UBC* reference gene at amplicon level to give semi-quantitative results for the first PCR. Different splice variants have been categorised according to *A. thaliana* classes in Fig. 3.1, for full details see Fig. 4.7.

These experiments confirmed distinct proximal and distal transcripts at *BrFLCA10*, with 3' and 5' polyadenylation sites respectively (Fig. 4.4D). Multiple termination sites (for both) were spread around the canonical *AtFLC COOLAIR* sites. Canonical, here, refers to the termination sites that were first annotated by Liu *et al*. (2010). Spread of polyadenylation sites has also been observed for *AtFLC* proximal transcripts; the majority terminated between -150 nt (upstream) and +50 nt (downstream) of the canonical site, but some terminated up to 300 nt upstream (within the small intron) (Zhe Wu, Dean lab, unpublished). *BrFLCA10* proximal polyadenylation sites were clustered in a ~100 bp window around the predominant site, slightly 5' of the canonical *AtFLC* proximal site. For the distal there is a greater spread of polyadenylation sites (-133 nt to +82 nt), with the predominant site also 5' of the canonical *AtFLC* distal site.

The prevalence of termination sites found upstream of the *AtFLC* canonical sites could reflect an experimental bias towards smaller transcripts when ligating the PCR products into a cloning vector (although that same bias should also have affected *Arabidopsis* results). *In planta* there may be greater representation of sites further downstream. Length of the distal transcript poly(A) tail, measured by number of adenosines (A), was highly variable: 9A – 144A, with an average length of 57A (1 StDev = 30). The opposite spreading patterns were observed in *A. thaliana*, with a smaller ~60 bp window around the canonical distal polyadenylation site and a larger window around the proximal site (Zhe Wu, Dean lab, unpublished). The greater spread of termination sites for the *AtFLC* proximal transcript, over the distal, was linked to the repressive chromatin state and slower RNA Pol II transcription associated with high proximal expression (Zhe Wu, personal communication). This does not explain why we see the opposite patterns for *BrFLCA10*.

4.2.4 **COOLAIR** *expression from the* **FLC5** *clade is uncertain*

The use of reverse gene-specific *FLC* antisense primers in the RT reaction should ensure that the cDNA is strand-specific, thereby preventing amplification of the sense transcript. Despite this, sense contamination (confirmed by sequencing) was found in some nested PCRs; for example, the larger band marked with 'S' in *BrFLCA3a* D16 (Fig. 4.4B). Non-strand specific amplification can occur at a low level during reverse transcription (Haddad *et al.*, 2007). This could be because of priming from other reaction products, such as structural regions of the input RNA or fragments of DNA from the DNAse treatment (Haddad *et al.*, 2007). These could effectively act as random primers for the reverse transcriptase, amplifying from either strand. Under normal PCR conditions, or where strand-specificity is not required, randomly primed products are probably at a level below

concern. It does, however, become a problem when using nested PCR reactions that are designed to amplify from low abundance amplicons in the initial PCR reaction. Non-strand specific amplification was confirmed possible under these reaction conditions by including controls without any RT primers in later experiments (-P) (Fig. 4.5).

It was still surprising to observe sense amplification because the *FLC* sense poly(A) site was thought to be 3' of the *COOLAIR* first exon (based on data from *AtFLC*), where the forward antisense primers annealed to. To check this, *BrFLCA10* and *BrFLCA3a* sense transcript 3'RACE experiments were performed (Fig. S5). This revealed a spread of poly(A) sites in *BrFLCA10,* some upstream and some downstream of the canonical *AtFLC* sense polyadenylation site. In contrast, *BrFLCA3a* sense termination was mostly at a single site slightly downstream of the canonical *AtFLC* site. In both examples, the *FLC* sense transcript can terminate downstream of the *AtFLC* canonical site in *COOLAIR* exon 1; this explains how sense transcripts can be amplified with the forward *COOLAIR* primers. Spread of polyadenylation sites is not surprising given what we observe for the antisense; it is possible that the *AtFLC* sense also uses a range of sites around that annotated. What is interesting is that the predominant *BrFLCA3a* termination site did not appear to have been polyadenylated. This could cause differences in processing and stability between the polyadenylated *BrFLCA10* mRNA and the non-polyadenylated *BrFLCA3a* mRNA, with the latter potentially more vulnerable to 3' exonucleases, less stable and less likely to be translated. Total RNA was enriched for polyadenylated transcripts prior to 3'RACE using PolyATtract® mRNA Isolation Systems (Promega), however, a small amount of contamination from non-polyadenylated transcripts can remain (see technical manual).

Accepting sense amplification as possible in the *COOLAIR* detection assays, this raises a question mark over the validity of the unspliced proximal Class I.ii transcript. It does not invalidate the other *COOLAIR* spliceoforms as splicing patterns and splice site orientation confirm they are antisense and not sense; Class I.ii, however, contains no introns and so could be either. 3'RACE products were not detected from *FLCA10* Class I.ii, suggesting that in this homologue the RT-PCR product represents sense contamination and/or unspliced total *COOLAIR* rather than proximal (Fig. 4.4D). In support of its conservation, however, amplification from positive samples was greater than amplification from the -P controls at *BoFLCC3b*, *BoFLCC9*, and *BoFLCC3a*, despite input cDNA being the same. This suggests some non-strand-specific contamination alongside antisense-specific amplification. Given these various interpretations, it cannot definitively be said that the Class I.ii transcript exists. This was the only transcript detected from *FLC5*, hence

COOLAIR expression from *FLC5* cannot be confirmed. *COOLAIR* may have been lost at this locus due to sequence divergence and change in function. Alternatively, *FLC5* may be expressed at too low a level to detect *COOLAIR*. Indeed, we observed low expression of both *FLCA3b* and *FLCC3b* in 12 *B. napus* cultivars (Fig. 4.2). Low expression of *FLCC3b* has previously been reported in *B. napus* and *B. oleracea* and it is thought to be a pseudogene (Pires *et al*., 2004; Okzaki *et al*., 2007; Razi *et al*., 2008; Zou *et al*., 2012).

4.2.5 **COOLAIR** *is differentially expressed between spring and winter* **B. napus**

So far, *COOLAIR* transcripts have been described from the A genome in a spring *B. rapa* (R018) cultivar and from the C genome in a semi-winter (DH1012) or winter (PSB) *B. oleracea* cultivar. If *COOLAIR* is transcribed from the *B. rapa* A genome and the *B. oleracea* C genome, it follows that it will also be transcribed in *B. napus*, where both genomes are present (although have diversified since hybridisation). The presence of *COOLAIR* was checked in *B. napus FLCA10*, *FLCA3a* and *FLCC2*: three copies of interest in terms of vernalization requirement and response. Two winter and two spring cultivars were selected to analyse differences in *COOLAIR* between the two.

The presence of total *COOLAIR* transcripts (proximal and distal exon 1) from *BnFLCA10* was confirmed in two spring (Stellar and Westar) and two winter (Express and Major) cultivars by semi-quantitative PCR (Fig. 4.6A). Total *BnFLCA10 COOLAIR* transcripts were expressed at a similar level in spring and winter types in the warm. Proximal and distal *COOLAIR* transcripts were then detected in *BnFLCC2* (Fig. 4.6B) and *BnFLCA3a* (Fig. S7), and proximal transcripts in *BnFLCA10* (Fig. S6). *BnFLCA10* distal products were not detected; whether because they are not expressed, or because the reaction conditions or primers were unsuitable is not known.

There are differences in *COOLAIR* splicing and expression patterns between spring and winter *B. napus* cultivars (Fig. 4.6B; Fig. S6; Fig. S7). Fewer or no distal transcripts were detected from *BnFLCC2* and *BnFLCA3a* spring versus winter cultivars, and fewer proximal transcripts were detected in *BnFLCA3a* spring versus winter cultivars. This could be because of lower expression of *BnFLCC2* and *BnFLCA3a* in Stellar and Westar compared with Express and Major (Fig. 4.2). A low *FLC* expression state is associated with low levels of both sense and antisense, but a relatively higher proximal:distal ratio (Liu *et al.*, 2010; Marquardt *et al*., 2014; Csorba *et al.*, 2014). Another possibility is that polymorphisms between spring and winter lines affect *COOLAIR* splicing, driving the production of one spliceoform over the other.

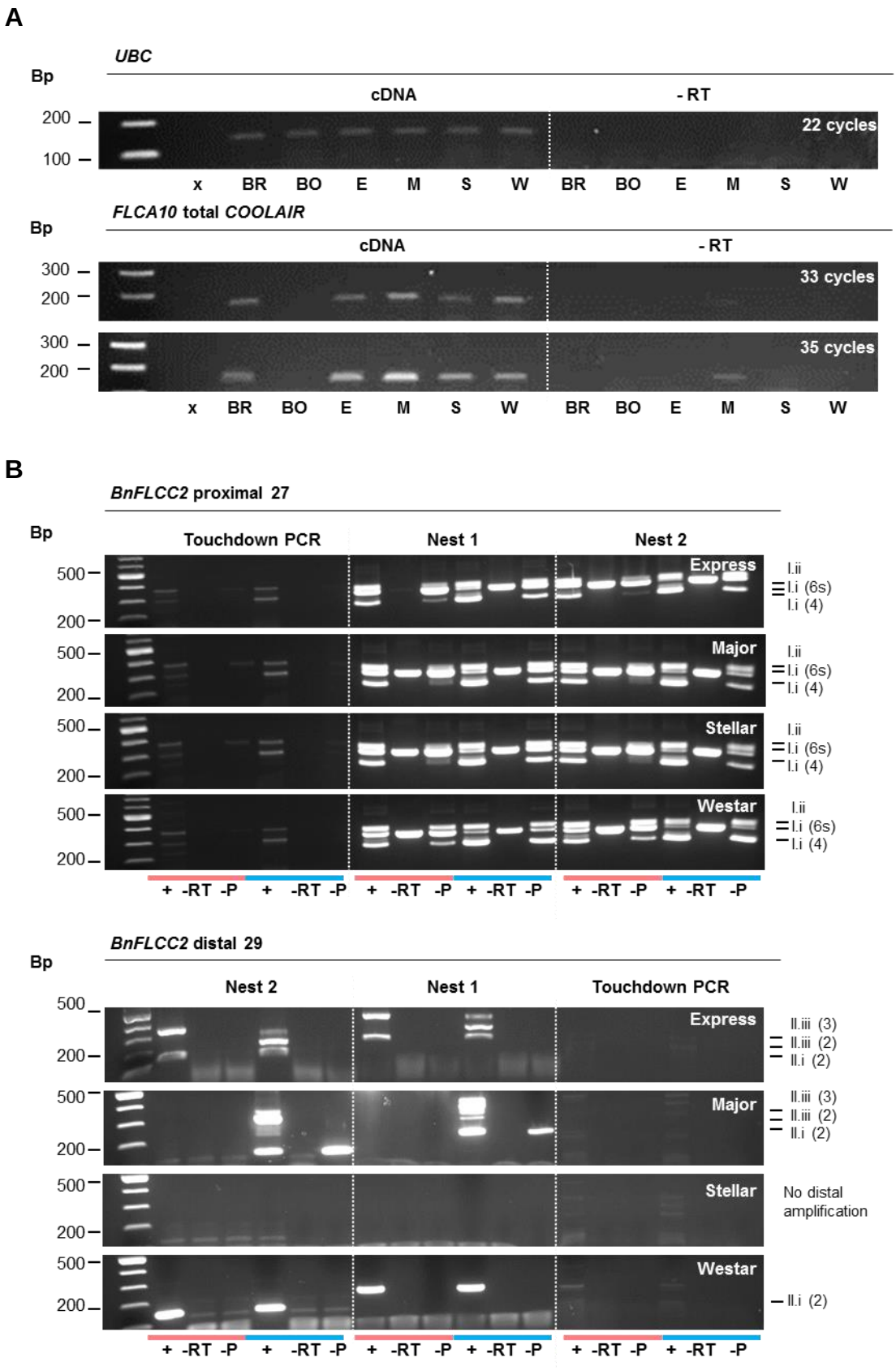

A
UBC
Bp
cDNA
- RT
200
100
22 cycles
x BR BO E M S W BR BO E M S W
FLCA10 total COOLAIR
Bp
cDNA
- RT
300
200
33 cycles
35 cycles
x BR BO E M S W BR BO E M S W
B
BnFLCC2 proximal 27
Bp
Touchdown PCR
Nest 1
Nest 2
500
200
Express
Major
Stellar
Westar
I.ii
I.i (6s)
I.i (4)
+ -RT -P + -RT -P + -RT -P + -RT -P + -RT -P + -RT -P
BnFLCC2 distal 29
Bp
Nest 2
Nest 1
Touchdown PCR
II.iii (3)
II.iii (2)
II.i (2)
No distal amplification
Key
X PCR negative control
+ cDNA sample
-RT No reverse transcriptase in RT reaction (control)
-P No primers in RT reaction (control)
Non-vernalized
Vernalized
BR B. rapa R018
BO B. oleracea DH1012
E B. napus Express
M B. napus Major
S B. napus Stellar
W B. napus Westar

Figure 4.6: Characterisation of *COOLAIR* in *B. napus* (previous page). (A) Semi-quantitative RT-PCR of total *BrFLCA10 COOLAIR* transcripts in non-vernalized winter (Major, Express) and spring (Stellar, Westar) *B. napus* cultivars, with amplicons first equalised for the *UBC* reference gene (top) at different cycle numbers (22 cycles shown). *B. rapa* R018 cDNA was a positive control, and *B. oleracea* DH1012 a negative control. (B) Touchdown nested RT-PCR experiments probed for *FLCC2 COOLAIR* proximal (top) and distal (bottom) transcripts from four *B. napus* cultivars, under non-vernalized (pink line) or two weeks vernalized (blue line) conditions. Initial touchdown PCR was followed by two nested PCRs (Nest 1 and Nest 2), where 27 and 29 refer to the proximal and distal primer sets, respectively. Input cDNA was equalised for the *UBC* reference gene at amplicon level to give semi-quantitative results for first PCR. Different splice variants have been categorised according to *A. thaliana* classes (Fig. 3.1), for full details see Fig. 4.7.

The spliceoforms that arise in *B. napus* are not necessarily the same as found in their *B. rapa* or *B. oleracea* progenitors. For example, Class II.iii (2,3) transcripts were transcribed from *BnFLCC2* winter lines, but not from a *BoFLCC2* winter line (PSB). This could be a consequence of polymorphisms that have arisen in *B. napus* after speciation. This would suggest that a relatively low number of polymorphisms at *FLC* (between *B. napus* and its progenitors or between cultivars) can significantly influence *COOLAIR* transcript architecture. Differences in *COOLAIR* expression patterns within the same line in ambient versus cold conditions were also observed in *B. rapa* and *B. oleracea* and are discussed in section 4.2.6 below.

4.2.6 There are many isoforms of COOLAIR in Brassica

Fig 4.7 summarises the *COOLAIR* transcripts detected under ambient or cold conditions across *Brassica* and other species in this study. Core gene architecture of the distal and proximal forms is maintained, primarily through conserved 5' splice sites and apparent conservation of 5' start and 3' polyadenylation sites. Conservation of 5' splice sites has been used as a guide for the division of *Brassica COOLAIR* transcripts into rough classes according to *AtFLC*; for example, if the distal intermediate exon terminates at the same site as in *Arabidopsis* (i.e. it retains the same 5' splice site for the second intron) then it is classified as Class II.ii or II.iii accordingly.

Fig 4.7 also reveals that within this basic architecture there is a great deal of variation. Where variation has arisen it is through changes in splicing: variation in the 3' splice site, incorporation of an extra exon, or removal of an extra intron. For example, the proximal Class I.i spliceoforms have been divided into types 1-6, with (4), (5) and (6) using a 3' intron splice site progressively more 5' of the canonical *A. thaliana* site in (1), thus shortening the intron (orange box in Fig. 4.7A). Alternate splicing of a lesser extent within types 1-6 are marked 's' (i.e. where there are alternate splice sites within a few nucleotides so that the length of the intron is not significantly altered). One interesting phenomenon is

the splicing of an extra intron within the canonical *COOLAIR* exon 1 in the Class I.i (3) transcript, present in both *FLCC9* and *FLCA10*. The sequence underlying *COOLAIR* exon 1 in *FLC1* contains a large insertion; this could be the reason that this new spliceoform has arisen.

Multiple proximal and distal spliceoforms, within the same or in different classes, can be transcribed from a single *FLC* homologue. These are not necessarily shared with their homeologues. For example, both *FLCA2* and *FLCC2* express the Class I.i (4) transcript in addition to a unique Class I.i transcript, (5) from *FLCA2* and (6) from *FLCC2* (orange box in Fig. 4.7A). Whereas *FLCA3a* solely expresses the distal Class II.i (1) transcript, *FLCC3a* additionally expresses Class II.ii (2) and (3) transcripts (blue and pink boxes in Fig. 4.7A). Such differences could be associated with a SNP at or close to a splice acceptor site, preventing its use. Up- or down-stream polymorphisms can also promote expression of different transcripts; for example, by creating a new splice acceptor site, or by altering splice site accessibility through changes in secondary structure or branch site sequence. An active versus repressive chromatin state affects RNA polymerase transcription rates and splice site selection through cotranscriptional coupling (Kornblihtt, 2007; Alló *et al.*, 2009; Nilsen and Graveley, 2010). Differences in splicing between homologues may therefore reflect respective expression levels.

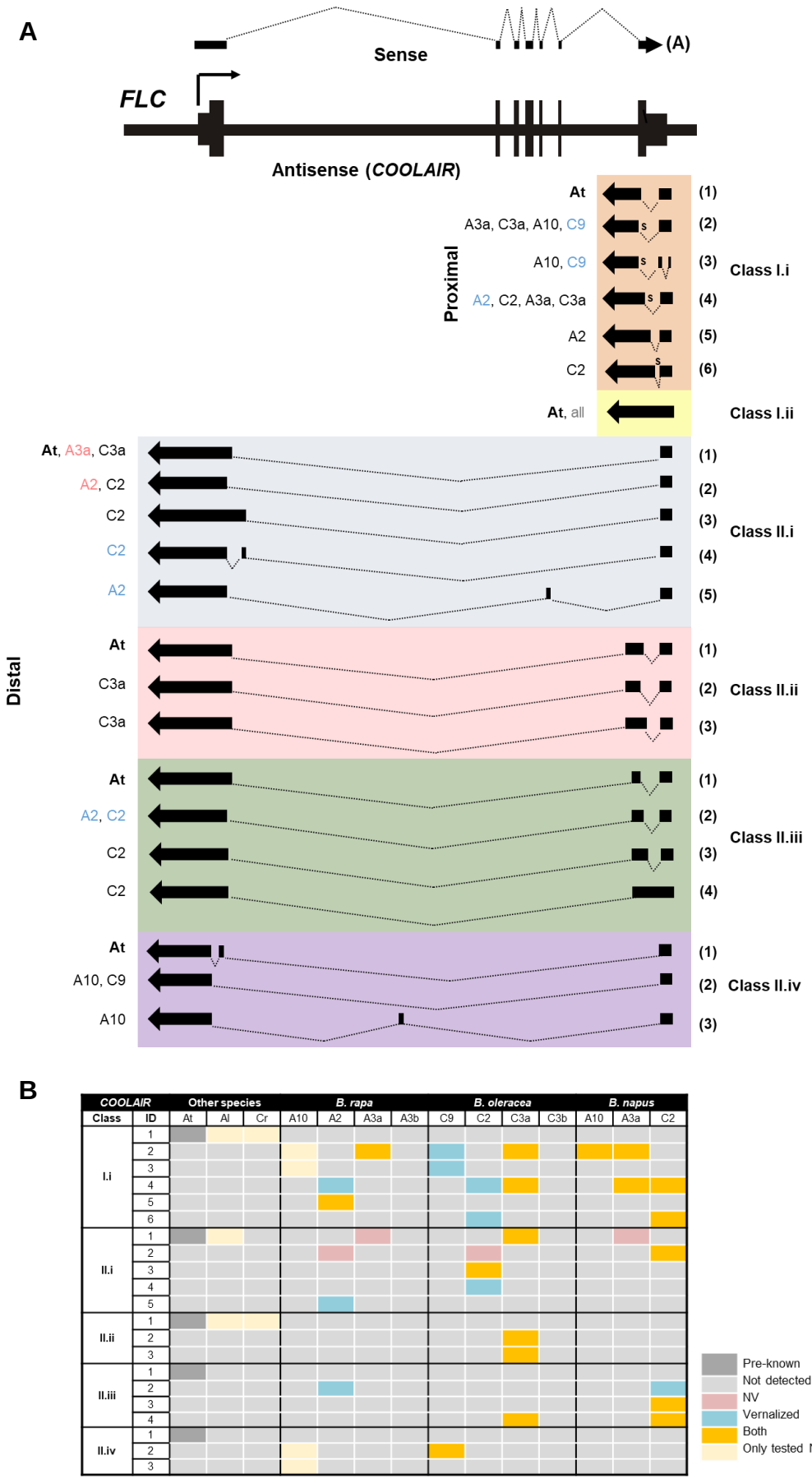


Figure 4.7: Summary of *COOLAIR* transcripts identified in *Brassica*. (A) Schematic representation of the *B. napus*, *B. rapa* and *B. oleracea COOLAIR* spliceoforms identified at each *FLC* locus. 'At' represents the *A. thaliana COOLAIR* classes (also see Fig. 3.1), according to which the *Brassica* transcripts have been classified. (B) Table of categorised Brassicaceae spliceoforms, with coloured boxes showing whether they were predominantly detected in the warm (pink), in the cold (blue), under both conditions (orange) or only tested in the warm (yellow). 'At' refers to *A. thaliana*, 'Al' to *A. lyrata* and 'Cr' to *C. rubella*.

One major difference between *FLC1*, *FLC2* and *FLC3* is their choice of 3' acceptor site for the large distal intron. Mutations at or adjacent to the canonical *AtFLC* Col-0 distal splice acceptor site has contributed to altered splicing of this intron, with *FLC3* using the canonical site, *FLC2* using a site 3' of this (or an alternate upstream site), and *FLC1* using the *AtFLC* Var2-6 site (which is conserved but not used in the other two homologues) (Fig. 4.8). The Var2-6 distal transcript sometimes incorporates an extra intermediate exon; this was also observed for *FLCA10* (Class II.iv (3)) but the exon is in a different position, further upstream of the *AtFLC* Var2-6 exon. This is interesting in light of the different flowering phenotypes driven by use of the Col-0 versus Var2-6 distal splice site in *A. thaliana* (Li *et al.*, 2015). In this case, we have three *COOLAIR* isoforms using three different splice sites within the same plant – it is interesting to speculate that each may have a different regulatory ability in consequence of this.

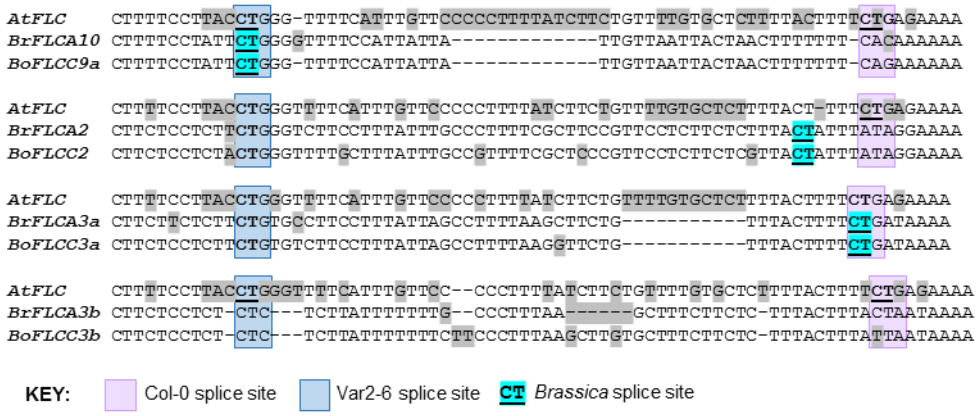


Figure 4.8: Distal *COOLAIR* large intron 3' acceptor splice sites in *Brassica FLC* homologues. Muscle multiple sequence alignment (Geneious R7) of A and C genome *FLC* homeologues with *AtFLC* Col-0. The *AtFLC* Col-0 and Var2-6 splice site positions are boxed in pink or blue, respectively. Splice sites used in the Brassica *FLC* homologues are highlighted in blue. The distal isoform has not been detected from *BrFLCA3b* or *BoFLCC3b*.

Some splicing isoforms are only present (or predominate) in the warm (non-vernalized conditions) and some only in the cold (vernalized conditions) (Fig. 4.7B). For example, in *B. oleracea*, the *FLCC2* Class II.i (2) form was detected only in the warm, the Class II.i (4) form only in the cold, and the Class II.i (3) form under both conditions. Four transcripts were predominantly detected in the warm and nine in the cold, with *FLCC2* and *FLCA2* producing the greatest number of cold-induced alternate transcripts. Ambient temperature or cold-induced transcripts come from the same *FLC* homologue and are therefore probably not caused by sequence polymorphisms (even if the cultivar is heterozygous for that locus transcripts are still environmentally-induced). The use of alternate transcripts may instead be related to changes in spliceosome efficiency or *FLC* activity. During cold *COOLAIR* is upregulated and *FLC* downregulated (in *Arabidopsis* at least); shifts in

chromatin state, RNA Pol II transcription rate and expression level could all impact antisense splicing (Nilsen and Graveley, 2010). *FLCA3a* is interesting because it expresses the distal transcript only in the warm in both *B. rapa* and *B. napus*. Perhaps proximal transcripts independently downregulate this homologue during the cold?

In *A. thaliana*, the ratio of proximal:distal *COOLAIR* transcripts increases during the cold (Csorba *et al.*, 2014). If this is true for all *FLC* homologues, then the abundance of cold-induced distal spliceoforms in *FLCA2* and *FLCC2* could be a consequence of canonical transcription becoming less critical. In the warm, differential splicing of the proximal and distal transcripts can directly impact sense expression levels (Marquardt *et al.*, 2014; Li *et al.*, 2015). It follows that some *COOLAIR* spliceoforms are better regulators of the sense than others. One hypothesis is that *COOLAIR* transcripts that are more effective regulators are favoured in the cold. More likely, differential splicing is a random consequence of differences in spliceosome and gene activity under cold versus warm conditions. Some *BrFLCA2* transcripts, for example, were upregulated in the cold, despite the R018 cultivar from which they were expressed having no vernalization requirement for flowering (Fig. 4.3). It would be interesting to analyse splice patterns in a vernalization-responsive cultivar and see how they differ. Changes in *COOLAIR* splicing and polyadenylation, whether random or not, could have consequences for sense regulation.

In summary, variation between isoforms – between homologues or within homologues under different conditions – has probably arisen as a consequence of changes in the underlying sequence or transcriptional context. Alternate *COOLAIR* spliceoforms are likely to have altered secondary structure and potentially different sense regulation capabilities. Where sequence polymorphisms have given rise to spliceoforms that affect *COOLAIR* function and consequently plant flowering time, these may be selected for or against over evolutionary time. Those that are currently transcribed in *Brassica* species may have a positive, negative or no effect on *COOLAIR* regulatory function. Expression of unique transcripts under different environmental conditions (cold versus warm) may be purposeful, or an indirect consequence of the changing environment. Either way, changes in *COOLAIR* splicing may alter regulatory ability, giving rise to different behaviour in the warm versus the cold.

4.2.7 Distal COOLAIR *secondary structure is conserved in* BrFLC *homologues but H4 length varies*

We previously determined the secondary structure of the *B. rapa FLCA3a* distal Class II.i (1) transcript through SHAPE-probing (Sanbonmatsu lab, see Chapter 3). Distinct *COOLAIR* spliceoforms and nucleotide sequences could give rise to alternate secondary structures in the other *Brassica FLC* homologues. The three *B. rapa FLC* homologues (with conserved *COOLAIR*) each express a unique distal transcript: Class II.i (1), *FLCA3a*; Class II.i (2), *FLCA2*; Class II.iv (2), *FLCA10* (Fig. 4.7). Additional distal isoforms are transcribed from *BrFLCA2* and *BrFLCA10*, but the above are most similar in genetic architecture – with two exons divided by one large intron – to the chemically-probed *AtFLC* Col-0 II.i secondary structure. These three transcripts each use an alternate 3' splice acceptor site for the large intron. *BrFLCA3a* uses the canonical *AtFLC* Col-0 splice acceptor site, whereas *BrFLCA2* uses a site 3' of this, and *BrFLCA10* uses the *AtFLC* Var2-6 splice site even further 3' (Fig. 4.8; Fig. 4.9B). In *A. thaliana*, we observe major structural changes between Col-0 and Var2-6 due to alternate use of these splice sites. These changes would be even more extreme without the extra exon in Var2-6. We predicted the distal secondary structures of the other *B. rapa FLC* homologues to observe how they are affected by diverse nucleotide sequence and differences in intron splicing.

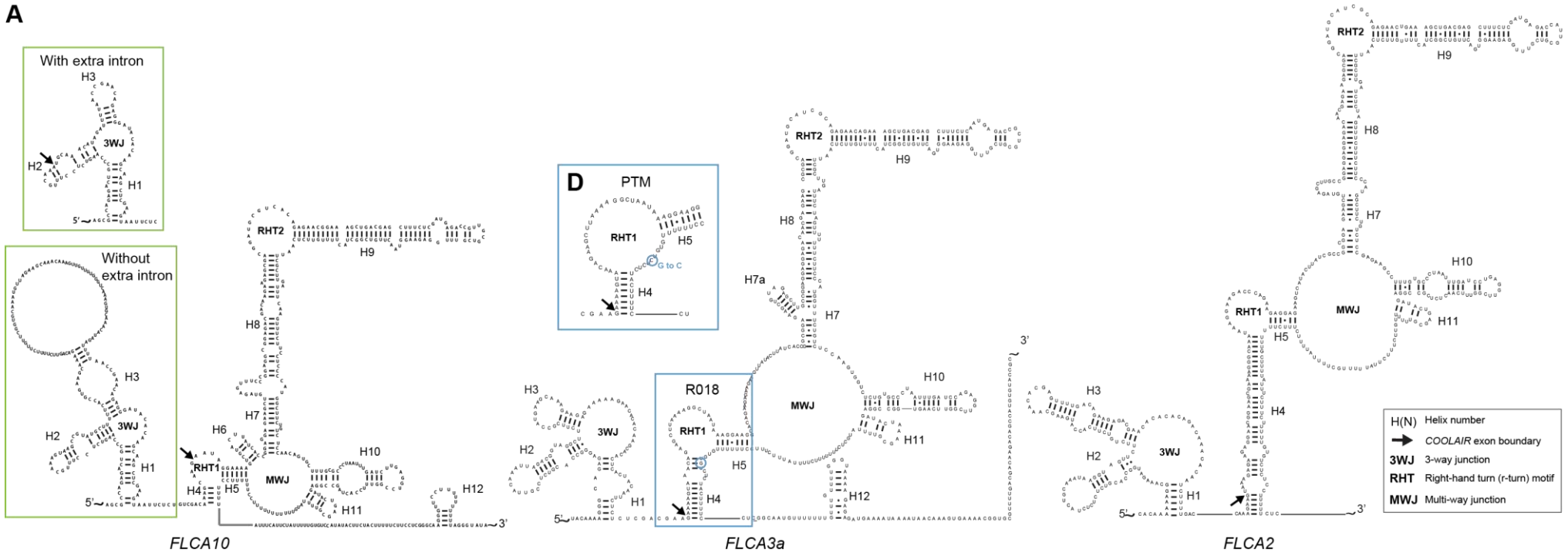
Distal secondary structures for the *BrFLCA10* and *BrFLCA2* homologues were predicted with reference to the SHAPE-probed structure of *BrFLCA3a*, using the method developed by Weinberg *et al.* (2007) (Fig. 4.9A). As for the five Brassicaceae species analysed in Chapter 3, global secondary structure of the distal *COOLAIR* has been conserved. They each contain two right hand-turn motifs, a multi-way (four- or five-way) junction in the second exon and a three-way junction in the first exon. Helices H7-H9 underlie the first sense exon and MADS box domain and have consequently been well-preserved.

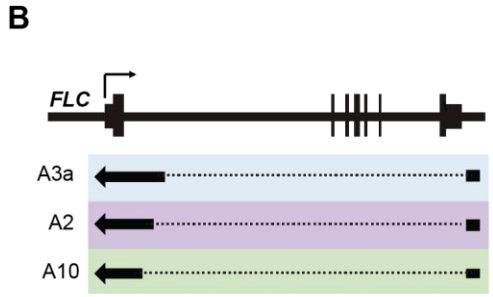
The three-way junction exists in all three but is less well conserved. Loss of the conserved 5' block has resulted in polymorphisms which alter the junction loop size and relative positions of the three helices. There are also changes to the size and complexity of H3, with an insertion in *BrFLCA10* giving rise to a large terminal loop. This may base pair within itself to form a more structurally complex H3. Interestingly, there is an alternate proximal Class I.i spliceoform in *BrFLCA10*, where a large part of this addition is spliced out as an extra intron, giving rise to the alternate exon 1 three-way junction (boxed above *BrFLCA10* in Fig. 4.9A), more like that seen in other species. Splicing of this first exon has only been observed in *BrFLCA10* and *BoFLCC9*. We therefore have a case where

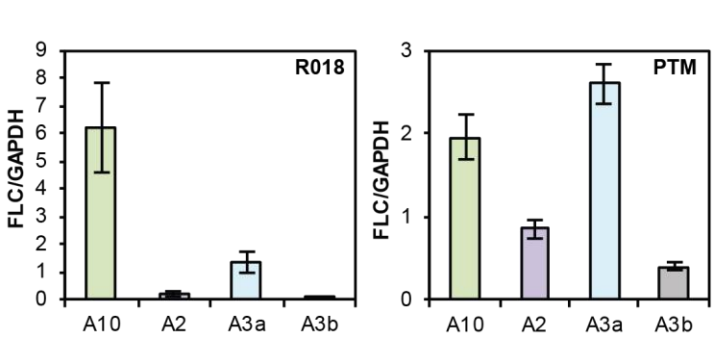
sequence polymorphisms (an indel) which cause significant change in RNA structure have been corrected for. We were unable to detect splicing of this extra intron in the distal transcript, although cannot rule out its existence. It may be that the *BrFLCA10* distal structure is already significantly disrupted (see below) and *COOLAIR* non-functional, hence there is no need to splice out this intron.

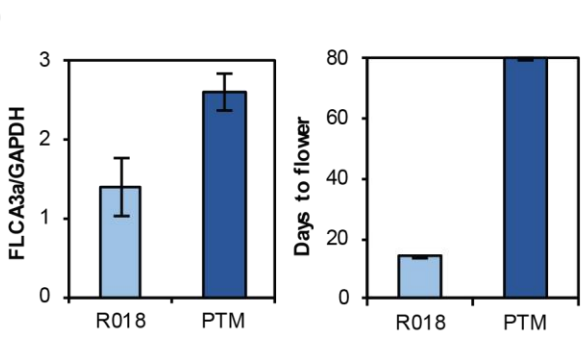
Aside from the changes in H3, the three structures differ in the length of helix H4. *BrFLCA10* uses the Var2-6 3' splice acceptor site and consequently is missing all of the sequence from the terminal exon that gives rise to the left-hand side of H4. Without any compensatory sequence (from an intermediate exon) it is predicted to form a very short 4 bp helix. This maintains the first right-hand turn motif. If these nucleotides do not base pair *in vivo* then *BrFLCA10* will lose the first r-turn motif altogether, severely disrupting secondary structure. *BrFLCA2* and *BrFLCA3a* use splice sites that are 5' of the Var2-6 site and consequently are able to form longer H4 helices (Fig. 4.8). Although the *BrFLCA3a* 3' acceptor site is the same as the canonical Col-0 site, and the *BrFLCA2* site is 3' of this, *BrFLCA2* is predicted to have a 29 bp H4, more like H4 in *AtFLC* Col-0 than the 11 bp H4 in *BrFLCA3a*. This is because *BrFLCA3a* contains an 11 bp deletion in the sequence pertaining to the left-hand side of H4 which removes nucleotides previously involved in base pairing (Fig. 4.8). We therefore have three *BrFLC* homologues each with a different length helix H4. In Chapter 3, we proposed that the length of H4 affected *COOLAIR* function, with a shorter helix less efficient at regulating the *FLC* sense. This was based on observations in *A. thaliana*, where function and structure of the Var2-6 and Col-0 *COOLAIR* transcripts were compared (Li *et al.*, 2015; Hawkes *et al.*, 2016). If this hypothesis is correct, then *BrFLCA10* with the shortest H4 would have the weakest *COOLAIR* (in terms of its ability to regulate *FLC*) and *BrFLCA2*, with the longest H4, the strongest.

Figure 4.9: Secondary structure and splicing of the distal *COOLAIR* at three *B. rapa FLC* loci (next page). (A) Predicted secondary structures of the major distal *COOLAIR* isoforms from *B. rapa* R018 *FLCA10* (Class II.iv (2), left) and *FLCA2* (Class II.i (2), right), with the SHAPE-probed secondary structure of *FLCA3a* Class II.i (1) in the centre. An alternate three-way junction (when a possible extra intron is spliced out) is given for the *FLCA10* distal structure (green boxed region). An alternate predicted H4 and r-turn is shown for the *FLCA3a* distal structure in the *B. rapa* PTM cultivar compared with R018 (blue boxed region), with the polymorphism between the two circled. (B) Schematic of the major *COOLAIR* distal isoforms from *FLCA10*, *FLCA2* and *FLCA3a*, revealing the relative positions of their 3' intron splice sites. (C) Expression of spliced sense *FLC* mRNA for the four homologues in *B. rapa* R018 spring (left) and PTM winter (right) cultivars under non-vernalized conditions, normalised to the *GAPDH* reference gene. Error bars are 1 St Dev. from mean for two (PTM) or three (R018) biological replicates. (D) Comparison of *FLCA3a* spliced expression levels for R018 and PTM (left) and non-vernalized flowering times, recorded as days to flower (right). PTM had not flowered by 80 days. Error bars represent the range in flowering times for three plants per cultivar.


A
With extra intron
Without extra intron
FLCA10
FLCA3a
FLCA2
D
PTM
R018
RHT1
RHT2
MWJ
3WJ
H(N) Helix number
COOLAIR exon boundary
3WJ 3-way junction
RHT Right-hand turn (r-turn) motif
MWJ Multi-way junction


B
FLC
A3a
A2
A10


C
R018
PTM
FLC/GAPDH
A10
A2
A3a
A3b


D
FLCA3a/GAPDH
Days to flower
R018
PTM

We present three unique *COOLAIR* isoforms, each associated with one *FLC* homologue in *B. rapa* and each with potentially different regulatory abilities. In contrast to the Col-0/Var2-6 story, this is a system where there are *COOLAIR* transcripts of ranging efficiency within a single plant. If each *COOLAIR* regulates its own *FLC* copy in *cis* then we would expect expression differences between homologues to be at least partially driven by differences in its *COOLAIR* isoform. In *B. rapa* R018, sense transcript expression from *FLCA10* is high, from *FLCA3a* intermediate and from *FLCA2* low (Fig. 4.9C). This correlates with the proposed regulatory abilities of their distal *COOLAIR* transcripts. *COOLAIR* at *BrFLCA10* has a short helix H4 and is therefore considered a poor regulator of the sense transcript; *COOLAIR* at *BrFLCA2* has a long H4 and is therefore considered a good regulator of the sense. Accordingly, the *BrFLCA10* sense transcript is strongly expressed whereas the *BrFLCA2* sense transcript is weakly expressed. *BrFLCA3a*, with the intermediate length H4, is expressed at an intermediate level. This is a relationship that holds true for the A genome across many of the *B. napus* cultivars analysed in Fig. 4.2, although not the winter *B. rapa* cultivar, Purple Top Milan (PTM) (Fig. 4.9C). There are multiple polymorphisms between *B. rapa* PTM and R018 that could contribute to differences in flowering time and *FLC* expression (Fig. 4.9C and D). Although *BrFLCA10* expression is higher in R018, the R018 allele contains the G to A SNP in intron six that gives rise to alternate sense splicing and non-functional proteins, hence flowering time is earlier (Yuan *et al*., 2009; R018 sequenced in Chapter 5). *BrFLCA3a* sense expression is higher in PTM than in R018 (Fig. 4.9D). Interestingly, the PTM allele contains a G to C SNP in helix H4 that prevent C-G base pairing, and is consequently predicted to shorten H4 to 8 bp and alter the first r-turn (Fig. 4.9D). It is interesting to speculate that the higher *FLCA3a* expression in PTM is at least partially contributed to by this SNP, which antagonises regulation of the sense transcript by the distal *COOLAIR*.

4.2.8 Distal COOLAIR *secondary structure is conserved between* FLC *homeologues*

Covariation analysis of predicted distal secondary structures for the *B. oleracea FLCC9a*, *FLCC2* and *FLCC3a* homeologues (against their A genome homeologues) reveal that polymorphisms have had little effect on global or local secondary structure (Fig. 4.10). The majority of SNPs maintain secondary structure because they are located at the end of a helix, adjacent to or in a loop or single-stranded region, or because there has been a covariant base pair mutation.

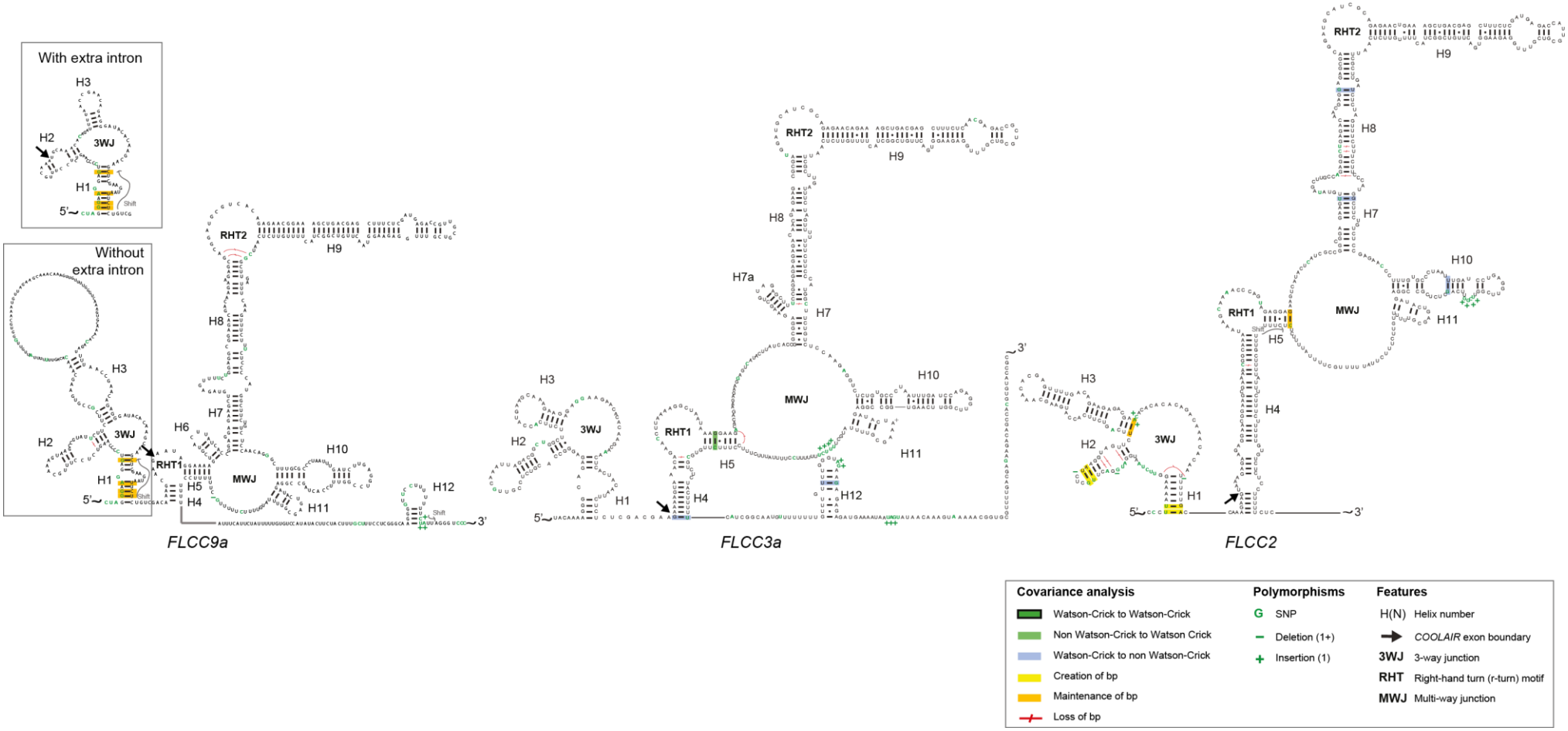


Figure 4.10: Secondary structure of the distal *COOLAIR* at three *B. oleracea FLC* loci. Predicted secondary structures of the major distal *COOLAIR* isoform from *B. oleracea* Alboglabra *FLCC9a* (Class II.iv (2), left), *FLCC3a* (Class II.i (1), middle), and *FLCC2* (Class II.i (2), right). An alternate three-way junction (when the possible extra intron is spliced out) is given for the *FLCC9a* distal structure (grey box). Covariance and polymorphism analyses were performed via sequence and structure comparison with their A genome *FLC* homeologues (Fig. 4.9).

There are also examples of new base pairs being created by SNPs (replacing lost base pairs elsewhere), such as in helix H2 of *BoFLCC2*, or proposed shifts in sequence to allow the maintenance of base pairs in that section, i.e. the shift of sequence by 1 bp to the right to maintain H5 in *BoFLCC2*. In *BoFLCC9a* a more dramatic shift of sequence is proposed to maintain base pairing in H1, despite multiple SNPs in the sequence pertaining to its left-hand side. Sequence shifts were also predicted to maintain structure in *A. alpina* H1, H4 and H5 and *C. rubella* H4, supporting their existence. Secondary structure of the major distal transcripts is predicted to be near-identical between *FLC* homeologues and shows signs of conservation; chemical-probing is required to validate this.

This contradicts the hypothesis that a long H4 is associated with strong *COOLAIR* repression [of *FLC*] because high *FLCC2* expression has been observed in *B. oleracea* (Irwin *et al*., 2016) and some *B. napus* cultivars (Fig. 4.2). If *COOLAIR* from *BrFLCA2* is a strong repressor, why isn't *COOLAIR* from *BoFLCC2* with its near-identical structure? This could be because the hypothesis is incorrect and *COOLAIR* regulation is unrelated to secondary structure or H4 length. Multiple other polymorphisms exist between *FLCA2* and *FLCC2*, however, that could impact sense expression levels through a non-antisense mechanism. Polymorphisms unique to *FLCC2* may drive its higher expression, overriding the effect of a strong *COOLAIR* regulator. Alternate forms of the distal transcript expressed from *BoFLCC2* (and not found in *BrFLCA2*) may be weaker repressors of the sense, counteracting and competing with the Class II.i (2) strong regulator. It is problematic to relate high or low expression levels of *FLC* to differences in *COOLAIR* when multiple other sequence polymorphisms exist between *FLC* copies. Some of these undoubtably play a role in determining *FLC* expression level and/or protein function. In Chapter 5, we therefore set out to determine how much, if at all, *COOLAIR* is contributing to differences in *FLC* expression and flowering time.

4.3 Discussion

4.3.1 Control of flowering time by FLC *homologues in polyploid species*

Polyploidy gives rise to genetic redundancy; duplicated genes may acquire deleterious mutations and be silenced or advantageous mutations and undergo neo- or subfunctionalization (Walsh, 1995). In cases where a dosage effect gives a selective advantage, all copies may evolve under purifying selection but with relaxed constraint (Lynch and Conery, 2000). This allows for greater variation in expression levels and consequently more intermediate phenotypes (Osbourn *et al.*, 2003). Original sequences of pseudogenes are rapidly eroded by mutations (Walsh, 1995); here, the eight analysed *Brassica FLC* loci all shows signs of sequence conservation. *Brassica* species therefore present an interesting alternative system to *Arabidopsis* with multiple copies of *FLC* within a single plant, each potentially contributing to flowering time. Previous studies suggest a conserved role for *FLCA10*, *FLCA3a*, *FLCC3a*, *FLCA2*, *FLCC2* and *FLCA3b* in flowering time control but not for *FLCC9* or *FLCC3b*; the latter contains a premature stop codon and is therefore thought to be non-functional or a recent pseudogene (Osborn *et al.*, 1996; Schranz *et al.*, 2002; Pires *et al.*, 2004; Lin *et al.*, 2005; Long *et al.*, 2007; Kim *et al.*, 2007; Lou *et al.*, 2007; Okazaki *et al.*, 2007; Razi *et al.*, 2008; Yuan *et al.*, 2009; Zhao *et al.*, 2010; Hou *et al*., 2012; Wu *et al.*, 2012A; Zou *et al.*, 2012; Xiao *et al.*, 2013; Ridge *et al.*, 2015; Irwin *et al.*, 2016; Schiessl *et al.*, 2017). This study reveals cumulative *FLC* expression levels to be higher in winter versus spring *B. napus* lines, supporting dosage-dependent regulation of flowering time by multiple copies.

We put forward *FLCA10*, *FLCA3a* and *FLCC2* as important players in the spring/winter difference, due to their high expression in winter and low expression in spring lines. This is dependent on greater production of functional protein in higher expression lines. There is support for *FLCA10* being important both in vernalization response (Osborn *et al.*, 1996) and in flowering time control (Long *et al.*, 2007; Okazaki *et al.*, 2007; Hou *et al*., 2012; Schiessl *et al.*, 2017). *B. rapa* R018 flowers rapidly without vernalization despite high expression of *FLCA10*. In this case, the R018 *FLCA10* allele has the A to G mutation at the 5’ splice site of intron six which gives rise to a non-functional protein (Yuan *et al.*, 2009). Major flowering time QTLs have also been reported at *FLCC2* (Okazaki *et al.*, 2007; Zou *et al.*, 2012). There is less evidence for *FLCA3a* being a major player (Zou *et al.*, 2012). We reveal that *BrFLCA3a* mRNA transcripts are poly(A) bimorphic, with the majority non-polyadenylated. Bimorphic poly(A) transcripts may arise from reduction and loss of the poly(A) tail after initial polyadenylation (Katinakis *et al.*, 1980). Without the protection

of a poly(A) tail, *BrFLCA3a* transcripts could be rapidly degraded by the nuclear exosome, reducing their influence on flowering time (Moore and Proudfoot, 2009). This may partially explain why *B. rapa* R018 is early flowering, despite high levels of *FLCA3a* mRNA. It would be useful to check whether or not *FLCA3a* protein levels are lower than predicted from mRNA expression levels.

FLCA3b, in contrast, was expressed at a low level in the six winter lines in this study, but has been linked to major flowering time QTLs in *B. rapa* (Schranz *et al.*, 2002) and *B. napus* (Zou *et al.*, 2012). Differential splicing of the *FLCA3b* sense transcript introduces a premature stop codon, which drives early flowering (Zou *et al.*, 2012). *BnFLCA3b* levels, although low in winter lines, were even lower in spring lines. We should therefore not ignore the potentially strong regulatory ability of the *FLCA3b* protein, despite low mRNA transcript levels. *FLCA3b* may be more potent because it is genomically close to *BnFRI-A3* (a potential upregulator) and the cold-responsive gene *BnCBF-A3*. Zou *et al.* (2012) hypothesised that these may have clustered together during translocation events and are now coexpressed and coregulated.

Defunct FLC proteins can contribute to early flowering in high *FLC* expression lines (Yuan *et al.*, 2009). In general, however, spring *Brassica* lines (in this study) exhibited lower *FLC* expression levels. Perhaps changes that work through sense expression levels offer more subtle and specific adaptations to environmental change than on/off protein functionality. *cis* non-coding mutations at *FLC* in *Arabidopsis* have been found to alter flowering time in the warm and after vernalization through direct or indirect changes in sense transcription rather than protein function (Shindo *et al.*, 2006; Coustham *et al.*, 2012; Li *et al.*, 2014; Li *et al.*, 2015; Qüesta *et al.*, 2016). Expression level variation can also arise from mutations in *trans*-regulators, such as *FRIGIDA* (Shindo *et al.*, 2005). In *Brassica*, there is even greater potential for variation in expression levels through dosage effects (Osborn *et al.*, 2003). This may explain the high preservation rate of *FLC* genes in *Brassica* (Lynch and Conery, 2000).

Whereas *BnFLCA10* and *BnFLCA3a* in two winter *B. napus* cultivars are stably downregulated after two or more weeks of cold, *BnFLCC2* is reactivated and flowering delayed. If this is consistent across winter lines, it suggests *BnFLCA10* is important for cold requirement but not response, and *BnFLCC2* for both. In this way, *BnFLCC2* is similar to the Lov-1 *AtFLC* allele which requires nine weeks of cold to fully saturate its vernalization requirement (Coustham *et al.*, 2012). Allelic variation at *FLCC2* in *B.*

oleracea confers differences in vernalization response and consequently heading date (Irwin *et al.*, 2016). *BnFLCC3a* also reactivated after cold and could be involved in vernalization response. Pires *et al.* (2004) associated early flowering with loss of *BnFLCC3a*. *BnFLCC9a* was highly expressed in both spring and winter lines in this study, with upregulation after cold treatment in flowering spring lines. *FLCC9* may have evolved a new role in floral development and/or seed set. Although rare, advantageous mutations leading to functional divergence are thought to have a much higher probability of being fixed than null mutations after gene duplication (Walsh, 1995).

Brassica napus is an allotetraploid, within which two functional genomes (A and C) have merged. Hybridisation events have the potential to change protein-regulator interactions and disrupt regulatory networks (Osborn *et al.*, 2003). Chromosomal rearrangements are frequent within the first few generations after polyploidisation, and are thought to play an on-going role in speciation (Gaeta *et al*., 2007; Cheng *et al*., 2013). The *B. napus* hybridisation event likely altered the floral regulatory network, with resultant copy number variation and sub-functionalisation of flowering time genes, such as *FLC*, contributing to the evolution of different morphotypes (Schiessl *et al*., 2017). In *B. rapa*, *FLCA10* has been mapped to QTL for vernalization response (Osborn *et al.*, 1996); whereas in *B. oleracea*, *FLCC2* has predominantly been mapped to early-late flowering time QTL (Okazaki *et al.*, 2007). This study suggests opposite roles in *B. napus*. We further note that a short period of cold is sufficient to switch off the A genome in *Brassica napus*, but a much longer period of cold is needed to switch off the C genome. These alternate roles may be a consequence of altered interactions post-hybridisation, with the C genome becoming responsible for cold response and the A genome for cold requirement.

Surprisingly, the C genome *FLC* homologues were reactivated to higher than ambient temperature expression levels after vernalization. To our knowledge, this has not previously been observed for reactivating *Arabidopsis FLC* alleles (Dean lab, personal communication). This suggests a level of repression has been removed (or activation gained) during cold treatment. We observed opposing expression patterns of A and C genome *FLC* homeologues in the two winter lines from the *B. napus* vernalization time-series. One hypothesis is that A genome *FLC* homologues directly or indirectly downregulate their C genome counterparts in normal (non-vernalized) conditions (or outcompete them for positive regulators). During cold, we propose *COOLAIR*-driven transcriptional downregulation of A and C genome *FLC* copies and PHD-PRC2 complex accumulation in the nucleation region, in an *Arabidopsis*-like mechanism (Hepworth and

Dean, 2015). After insufficient cold, lack of PHD-PRC2 spreading causes reactivation of C genome *FLC* copies, whereas spreading at A genome *FLC* copies epigenetically silences them. The C genome *FLC* copies are reactivated to above their non-vernalized levels, due to the loss of A genome *FLC* repression. Cross-talk between *FLC* homeologues from the A and C genomes could have arisen to maintain appropriate levels of floral repression after *FLC* dosage doubled during hybridisation.

If there is interaction between homeologues in *B. napus*, we cannot rule out interaction between homologues in the diploid *B. rapa* or *B. oleracea*. It is tempting to speculate, however, that this is a unique system that has evolved due to hybridisation of two distinct genomes in *B. napus* and is only possible (or necessary) where we have two *FLC* copies with near-identical sequence. This is supported by evidence that recombination is more frequent at large regions of homoeology between A and C chromosomes (Gaeta *et al*., 2007). Structural, functional, and epigenetic cross-talk between homeologues has been established (Song *et al*., 1995; Chalhoub *et al*., 2014). One possibility is that sense transcripts from the two homeologues could base pair, triggering an RNAi-like mechanism of degradation. This would potentially reduce transcript number from both homeologues, unless A genome *FLC* expression is higher to begin with. Another possibility is the physical interaction of the two chromosomes, perhaps with one blocking access of transcriptional machinery or *cis*-regulatory proteins to the other. Physical interaction can occur during G1 and S phases of the cell cycle, with homeologous exchanges and nonreciprocal transpositions, ranging from large segments to single SNPs, reported between the A and C subgenomes (Gaeta *et al*., 2014; Chalhoub *et al*., 2014). Interestingly, Chalhoub *et al.* (2014) noted 1.3 times more single nucleotide conversions from the A to the C genome than the reverse. Given their high sequence identity, the two homeologues may also compete for the same regulatory proteins. In section 4.3.3, we develop a further hypothesis: *COOLAIR* transcripts from A genome *FLC* copies not only regulate their own copy in *cis* but their C genome homeologue in *trans*.

Why would the A genome *FLC* copies regulate the C genome *FLC* copies, and not the reverse? It has been observed in plants that when two genomes merge, one subgenome outcompetes the other in terms of gene expression and activity (Schnable *et al*., 2011; Woodhouse *et al.*, 2014). This is evident through biased gene loss in the less dominant subgenome (Schnable *et al*., 2011). In the diploid *Brassica* species, subgenome III from the ancient triplication event was found to be more highly expressed and better conserved (less fractionated) than subgenomes I or II (Wang et al., 2011). In *B. napus*, there is

absence of significant bias towards either subgenome, with patterns of gene dominance or equivalence found in both (Chalhoub *et al*., 2014). Marginal biases towards C genome loss through nonreciprocal transpositions have been reported, however (Gaeta *et al*., 2007). Schiessl *et al.* (2017) further noted that copy number variation in flowering time genes *FLC*, *PHYTOCHROME A* and *GIBBERELLIN-3-OXIDASE 1* involved duplications in the A genome and corresponding losses in the C genome.

4.3.2 Expression of COOLAIR *from* FLC *homologues in polyploid species*

At least six of the ten *FLC* copies in *B. napus* are predicted to contribute to overall flowering time, with some having distinct roles in this process. *COOLAIR* has been conserved at three of the four *FLC* loci in *B. rapa* and *B. oleracea*, and may also be expressed from the fourth (*FLC5*), but at levels undetectable by the developed assay. *COOLAIR* transcripts were further detected in three out of three tested *B. napus* paralogues: *FLCA10*, *FLCA3a* and *FLCC2*. This confirms that *COOLAIR* has been conserved at multiple *FLC* loci within a plant, contrary to earlier reports (Li *et al.*, 2016). Despite sequence divergence between *FLC* homologues, *COOLAIR* lncRNAs are highly conserved in terms of gene architecture, start and termination sites, 5' splice sites and secondary structure. Strong 5' splice sites are thought to have coevolved with transcripts with low intron numbers [such as *COOLAIR*] across eukaryotic genomes (Irimia *et al*., 2007). Strong conservation of 5' splice acceptor sites allows division of *COOLAIR* into classes roughly similar to that of *Arabidopsis*. Their persistence supports a role in sense regulation and flowering time control in *Brassica*. Although there is conservation of global transcript architecture and key structural domains, 3' splice site and exon/intron inclusion or exclusion vary. A greater number of *COOLAIR* variants may have evolved in *Brassica* (than in *Arabidopsis*) due to rapid divergence after gene duplication (Song *et al*., 1995; Osborn *et al.*, 2003; Gaeta *et al.*, 2007; Cheng *et al*., 2013), accelerated by the lower sequence constraint characteristic of lncRNAs (Pang *et al.*, 2006; Kutter *et al.*, 2012; Nitsche *et al.*, 2015). If *COOLAIR* acts as an evolutionary tool to fine-tune flowering time in *Arabidopsis*, then multiple variants in *Brassica* could achieve greater precision in a more complicated system.

Multiple splicing isoforms of *COOLAIR* were identified within and between *FLC* homologues. Some were specific to winter or spring cultivars in *B. napus*. New spliceoforms can arise from mutations at or around splice sites, or from sequence polymorphisms in adjacent introns (Meyer *et al.*, 2011). Changes in relative size or sequence of adjacent exons can also alter selection of 3' splice site (Reed and Maniatis,

1986). Choice of splice site will be affected by local chromatin state due to kinetic coupling of RNA Pol II elongation rate and splicing (de le Mata *et al.*, 2003; Alló *et al.*, 2009; Moore and Proudfoot, 2009). Differences in splicing before and after cold were also apparent, i.e. there was an abundance of cold-induced *BoFLCC2* and *BrFLCA2 COOLAIR* transcripts. The spliced proximal isoform was only detected from *BoFLCC2* after cold. This is interesting because this is the form associated with downregulation of the sense transcript in *Arabidopsis* (Marquardt *et al.*, 2014). Several factors may contribute to alternative splicing in the cold. If *Arabidopsis* expression patterns are conserved, then *Brassica COOLAIR* transcripts will be upregulated in the cold. Hence, splice variants expressed at relatively low levels before vernalization can now be detected. Cold treatment may also alter splicing efficiency. Greater intron retainment has been shown to correlate with increased environmental stresses, such as cold (Ner-Gaon *et al.*, 2004). Shorter *FLC* sense splice variants were detected after vernalization in the *FLCA* and *FLCB A. thaliana* haplotypes (Caicedo *et al.*, 2004). Lower temperatures could also influence secondary structure of the unspliced transcript, affecting splice site selection (Meyer *et al.*, 2011). Changes in RNA processing may therefore result from cold-induced changes in gene expression, RNA Pol II elongation rate, splicing efficiency, or RNA secondary structure. Whether temperature-specific transcripts are actively regulated or passively induced, it is interesting that different *COOLAIR* isoforms – with potentially different functional capacities – exist under different environmental conditions.

Piecing together structural and architectural data, we build an understanding of variation and conservation at the distal *COOLAIR* in *Brassica*. Transcript architecture of the major distal form is preserved, with syntenic transcription of *COOLAIR* exons 1 and 2 divided by a large intron. The major structural domains – the 3-way junction at the 5’ end, the two right-hand turn motifs, the multi-way junction, and helices H7-H9 – are conserved, giving rise to similar global structures. Conservation of the latter is consequent of low DNA sequence variation (H7-H9 overlap the sense transcript). Presence of the other motifs, necessary for maintaining global structure, in non-protein-coding regions is revealing. This suggests that the basic structure is essential for *COOLAIR* function, with the position of exon 1 and the position of the proximal and distal polyadenylation sites critical. Other *COOLAIR* domains are subject to greater variation. Structurally, the length of helix H4, the length of the linker region between H1 and H4 (longer in *FLC2* than in *FLC1* or *FLC3*, Table S1), the size of the multi-way junction, and presence and complexity of H12 and the 3’ region can all vary (Fig. 3.11). This could suggest that these domains are less important

for *COOLAIR* function and hence have not been conserved. Alternatively, they may be subject to positive selection, with advantageous mutations being selected for.

One driver of structural variation is alternate splicing of the distal terminal exon. *FLC3*, *FLC2* and *FLC1* each use a unique progressively downstream distal 3' splice acceptor site (shortening the distal terminal exon). *FLC3* uses the canonical Col-0 distal splice site, *FLC1* uses the Var2-6 splice site, and *FLC2* an intermediate site. In *Arabidopsis*, a G to T SNP adjacent to the canonical CT splice site promotes use of the downstream Var2-6 site (Li *et al.*, 2015). In *FLC1* there is a C to T SNP within the CT splice site itself, in addition to an adjacent G to C SNP. This is an exciting example of the same natural variation in *COOLAIR* independently arising in *Arabidopsis* and *Brassica*. In *Arabidopsis*, the Var2-6 SNP not only alters distal transcript architecture but secondary structure, with the truncated second exon shortening helix H4. We hypothesised that the intermediate length H4 in Var2-6 was responsible for altered *COOLAIR* regulation of the sense and consequently the late flowering phenotype (Hawkes *et al.*, 2016). In *Brassica*, the three alternate splice sites at *FLC1*, *FLC2* and *FLC3*, in addition to other sequence polymorphisms, gives rise to three unique distal *COOLAIR* structures with different H4 lengths within a single plant. Due to an 11 bp deletion in *FLC3*, it is the *FLC2 COOLAIR* which has the longest H4 and is most similar to Col-0. We propose that the long H4 in *FLC2*, the intermediate H4 in *FLC3* and the short H4 in *FLC1* will result in distal *COOLAIR* transcripts with strong, intermediate and weak regulatory abilities, respectively. If each regulates its own *FLC* sense transcript in *cis*, then these differences may contribute to the differences in *FLC* expression levels between homologues. In support of this, *FLCA10* expression is highest, *FLCA3a* intermediate and *FLCA2* lowest in the *B. rapa* spring cultivar, R018, and the six analysed *B. napus* winter cultivars. In a winter *B. rapa* cultivar, PTM, *FLCA3a* expression is higher than in R018. This correlates with a SNP predicted to disrupt distal secondary structure, giving rise to a shorter helix H4. We do not yet know if this SNP is causative for the expression difference.

The global secondary structure of the distal *COOLAIR* in *Brassica* has withstood evolutionary selection, despite low sequence similarity, while variation in other regions – such as H4 – has potentially allowed adaptation to a changing environment. Natural variation at *COOLAIR* in *Arabidopsis* gave rise to a unique flowering time phenotype in Var2-6 (Li *et al.*, 2015). It is likely that some of the differences between *Brassica* spliceoforms will impact secondary structure and consequently sense regulation. This diversity may be utilised under different environmental conditions, such as ambient

temperatures versus cold, via the upregulation of specific transcripts. These examples showcase the abundance of splice variants and *COOLAIR* forms possible when the naturally rapid divergence of long non-coding RNAs is accelerated in a polyploid system. We know that there are multiple polymorphisms between and within copies and many of these will work through mechanisms other than *COOLAIR*. It is difficult to separate regulation by *COOLAIR* from other changes at *FLC* or in *trans*-regulators. Therefore, in Chapter 5, we isolate individual *Brassica FLC* copies and test the effect of mutations in *COOLAIR* on sense regulation.

4.3.3 COOLAIR *regulation of* FLC *in polyploid species*

We propose that in a system with multiple copies of *FLC*, each is regulated by its own *COOLAIR* in *cis*. This is supported by evidence for localised cotranscriptional regulation of *FLC* in *A. thaliana*, in addition to time and space restrictions in the nucleus. Transcriptional circuitry has been reported between sense and antisense transcripts at *FLC*, with expression levels tightly intertwined (Swiezewski *et al.*, 2009; Marquardt *et al.*, 2014; Wang *et al.*, 2014A). Single-molecule (sm)RNA FISH revealed that *COOLAIR* transcripts localise to the *FLC* locus at the chromosome from which they were (presumably) transcribed in the nucleus (Rosa *et al.*, 2016). In support of this, ChIRP experiments found that *COOLAIR* RNA associates with *FLC* genomic DNA, and is enriched at the nucleation and 3' regions (Csorba *et al.*, 2014). Li *et al.* (2016) proposed a master *COOLAIR* regulator transcribed from *FLCA2* in *B. rapa*, but this was before *COOLAIR* transcription from other copies had been detected by Hawkes *et al.* (2016) and in the current study. The conservation of *COOLAIR* at other *FLC* homologues supports an additional role for these in *FLC* sense regulation. Post-transcriptional regulation of *FLC* loci in *trans*, potentially in addition to *cis*-regulation, should be explored further. One tantalising possibility is that high sequence homology between *B. napus FLC* homeologues permits the coregulation of these by *COOLAIR* from either or both copies, not possible in the diploid *Brassica* progenitors.

We propose the *FLCA10* distal *COOLAIR* is a weak regulator of *FLCA10*, *FLCA3a* an intermediate regulator and *FLCA2* a strong regulator. This supports the hypothesis that *COOLAIR* transcripts from *FLCA2* are important flowering time regulators in *B. rapa* (Li *et al.*, 2016). A tentative proposal from our data is that A genome *FLC* copies can repress C genome *FLC* copies before cold. *COOLAIR* RNAs are one mechanism through which this could work. Numerous studies have shown that lncRNAs can regulate other genetic loci in *trans*; for example, the lncRNA *HOTAIR* silences *HOXD* and other genomic loci in

trans by the recruitment of PRC2 and LSD1 complexes (Rinn *et al.*, 2007; Tsai *et al.*, 2011). High sequence homology would facilitate *COOLAIR* binding to sense transcripts or genomic DNA at its homeologue as well as in *cis*. Considering hypothetical *COOLAIR* strengths, distal *FLCA2* transcripts are proposed to be strong repressors, but *FLCC2* sense expression is higher than *FLCA2* in some winter *B. napus* lines. This does not support C genome regulation by the A genome *COOLAIR*, although higher *BnFLCC2* expression could be a consequence of other factors. The *BnFLCA10* distal is proposed to be a weak repressor, it is therefore unlikely to efficiently downregulate either itself or *BnFLCC9*, accordingly we see high expression of both pre-vernalization. *BnFLCC9a* expression still increases to above pre-vernalization levels after cold; however, this is apparent only in spring lines, suggesting independent upregulation of *BnFLCC9a* during flowering. The *BnFLCA3a* distal transcript is proposed as an intermediate repressor, regulating both itself and *BnFLCC3a*. A further possibility is that this cross-talk works in both directions, with *COOLAIR* RNAs from both homeologues suppressing themselves and each other, thus maintaining a level of FLC protein more like that of their diploid progenitors.

COOLAIR expression has been maintained, despite high sequence diversification, at three of the four *Brassica FLC* clades. *COOLAIR* may have been lost, along with the functional protein, at *BoFLCC3b*, and is undetectable from the minimally expressed *BrFLCA3b*. Differences between the four *FLC* clades have likely been selected for over evolutionary time, with each developing distinct expression patterns, potencies and functions in flowering time control. Simultaneously, the *COOLAIR* lncRNAs at each locus would have diverged from each other. It is likely that polymorphisms at *COOLAIR*, arising more frequently in the non-coding than protein-coding transcripts, would have helped drive differences between *FLC* copies. The unique *COOLAIR* isoforms that have arisen have largely been conserved between *FLC* homeologues in *B. napus*, supporting distinct functional roles. Although this partly reflects stronger sequence homology between homeologues, we note that where there are polymorphisms they predominantly arise in single-stranded or loop regions. More recent mutations (post-speciation, within the last 10 Ka) at *FLC* and other flowering time genes in *B. napus* are thought to have driven major differences between morphotypes (Schiessl et al., 2014). We propose that some of these may work through changes in *COOLAIR*.

5 The effect of natural and induced variation at *COOLAIR* on flowering time

5.1 Introduction

The timing of flowering determines a plant's reproductive success and is therefore an important adaptive trait. Flowering time adaption has been extensively studied in *Arabidopsis thaliana*. *Arabidopsis* accessions grow over a wide latitudinal range, with natural variation in flowering time reflecting strong selective pressure to adapt to a new environment (Hepworth and Dean, 2016). Adaptations are complex and can reflect large or small-scale environmental change; early reports suggested a latitudinal cline in flowering time (Caicedo *et al*., 2004; Stinchcombe *et al.*, 2004), but a local response seems more likely (Johanson *et al.*, 2000; Shindo *et al.*, 2005; Shindo *et al.*, 2006; Ågren *et al*., 2017). In *A. thaliana*, a network of pathways converge on a set of common targets to quantitatively regulate genes involved in the switch from vegetative to floral state. Natural variation in flowering time has been mapped to a few major-effect QTLs, including *FLM*, *HUA2*, *VIN3*, *FRI* and *FLC* (Werner *et al.*, 2005A and 2005B; Shindo *et al.*, 2006; Strange *et al.*, 2011; Sánchez-Bermejo *et al*., 2012; Ågren *et al*., 2017).

Some *Arabidopsis* accessions complete their life-cycle rapidly whilst others remain vegetative over winter and flower in the favourable conditions of spring. Many flowering time adaptions work through changes to plant vernalization requirement (switching between rapid-cycling or winter habits) or response (differences in overwintering). Loss or gain of vernalization requirement has frequently been associated with changes at *FRI* (Johanson *et al.*, 2000; Shindo *et al.*, 2005; Strange *et al.*, 2011; Sánchez-Bermejo *et al*., 2012). Multiple loss-of-function *FRI* alleles have evolved independently over time to confer early flowering phenotypes (Le Corre *et al.*, 2002; Gazzani *et al.*, 2003; Shindo *et al.*, 2005; Werner *et al.*, 2005A). *FLC* loss-of-function alleles also eliminate vernalization requirement, but are less common in naturally occurring populations (Johanson *et al.*, 2000; Werner *et al.*, 2005A). Plants with vernalization requirement still exhibit a wide range of flowering times, largely due to epigenetic differences in their vernalization response (Shindo *et al.*, 2006). Major effect QTL underlying variation in vernalization response have been mapped to *FLC* (Strange *et al.*, 2011). Variation in vernalization response is indicated by the length of cold required to satisfy vernalization requirement, and correlates both with initial *FLC* expression levels (Sheldon *et al.*, 2000) and the stability of *FLC* repression (Shindo *et al.*, 2005; Shindo *et al.*, 2006).

cis polymorphisms at *FLC* can account for differences in vernalization requirement and response (Coustham *et al.*, 2012; Li *et al.*, 2014; Li *et al.*, 2015; Sánchez-Bermejo *et al*., 2012; Méndez-Vigo *et al.*, 2016; Qüesta *et al.*, 2016). Variation in the coding sequence can be limited; for example, there is 100% amino acid identity between the active Col-0 and C24 and the weak Landsberg *erecta FLC* alleles (Sheldon *et al.*, 2000). Instead, variation is primarily in non-coding regions, where key regulatory elements are found. Analysis of 1307 accessions revealed five major *AtFLC* haplotype groups with distinct non-coding sequence polymorphisms, *FLC* expression levels and rates of epigenetic silencing (Li *et al.*, 2014). For example, the *FLC* allele found in Lov-1, and other late flowering accessions within haplotype group 9, requires a long (nine week) period of cold to satisfy its vernalization requirement and become epigenetically silenced (Coustham *et al.*, 2012). This trait was mapped to four non-coding *cis* SNPs and linked to reduction in H3K27me3 [repressive chromatin marks] before and after vernalization at this locus (Coustham *et al.*, 2012). A C to T SNP in the first of a pair of RY elements in the nucleation region of *FLC* in the *vrn8* mutant prevents epigenetic silencing of *FLC* during cold by inhibiting VAL1 protein-binding via its B3 domain thus blocking PHD-PRC2 nucleation (Qüesta *et al.*, 2016). The latter was identified through a forward genetic screen, but demonstrates the significance of mutations in non-coding regions of *FLC*. Natural variation in *Arabidopsis* accessions can also work through changes in *COOLAIR* regulation of the sense transcript. A SNP which alters *COOLAIR* distal splicing is responsible for the late flowering phenotype of Var2-6 and other members of haplotype group 11 (Li *et al.*, 2015). Interestingly, two of the four causative Lov-1 SNPs overlap with the 3' end of the *COOLAIR* distal terminal exon (Coustham *et al.*, 2012). The authors speculated that this could drive the (observed) higher levels of distal polyadenylation (versus Col *FRI*) during cold and (via an unknown mechanism) antagonize epigenetic silencing.

Selective breeding in *Brassica* crops has led to the development of winter cultivars with, and spring cultivars without, vernalization requirement. Genetic analysis of breeding lines has focussed primarily on mutations that have a strong effect on flowering phenotype through changes to the FLC protein. Polymorphisms which alter sense transcript splicing to give non-functional proteins of *BrFLCA10* (Yuan *et al.*, 2009) and *BrFLCA2* (Wu *et al.*, 2012A) have been identified in early flowering lines. More recently, *cis* polymorphisms between two *FLCC2* alleles were found to confer differences in vernalization response and consequently heading date in *Brassica oleracea*, with 43 candidate SNPS from non-coding intronic regions (Irwin *et al.*, 2016). If regulation of *FLC* in *Brassica* is not unlike that in *Arabidopsis*, the latter can act as a useful model for the identification and characterisation

of important non-coding polymorphisms in the former. In Chapter 4, we observed parallels between *COOLAIR* distal splicing patterns in *Arabidopsis FLC* haplotype groups and *Brassica FLC* clades.

Multiple copies of *FLC* within *Brassica* plants, each with the potential to delay flowering, provide an extensive system in which to investigate non-coding flowering time adaptations. *COOLAIR* is transcribed from at least three of the four *FLC* homologues (*FLC1*, *FLC2* and *FLC3*) in *B. rapa* and *B. oleracea*. An abundance of different *COOLAIR* forms were detected within and across *Brassica* species, with natural variation in splicing and structure. *Brassica* species therefore provide an ideal model to test whether natural variation in flowering time has arisen through modifications to *COOLAIR* once (Var2-6) or multiple times over evolutionary history. This work further aims to understand how *COOLAIR* functions in polyploid or paleopolyploid species, and to test how far distinct *COOLAIR* isoforms contribute to differences between their respective *FLC* homologues.

5.2 Results

5.2.1 Natural cis *polymorphisms at* FLC *affect* COOLAIR *structure and processing*

17 winter and 13 summer *Brassica napus* cultivars from the JIC core diversity set (OREGIN; Table 2.2 and 2.3), and 15 winter cultivars from a Bayer CropScience (BCS) breeding set (Table 2.2) were selected and key regions – pertaining to *COOLAIR* – of three *FLC* copies (*BnFLCA10*, *BnFLCA3a* and *BnFLCC2*) sequenced (see Fig. 2.1 for method). These homologues were selected for their proposed role in flowering time and vernalization response and for their interesting *COOLAIR* isoforms (Chapter 4). Nucleotide sequences were aligned with available sequence from the PPS spring line (proprietary BCS) and divided into nominal alleles based on their sequence polymorphisms (Fig. 5.1-5.3). The frequency of alleles found in spring versus winter lines was used to infer whether they were more likely to confer a late or early flowering phenotype. Differences in vernalization response were more difficult to attribute to individual alleles. Any inferences are complicated by the recognition that different combinations of *FLC* alleles will give rise to different phenotypes, as will variation in other genes in the flowering time pathway.

There is almost no variation across winter cultivars at *BnFLCA10*, suggesting (1) that this allele confers late flowering and (2) that there is a higher mutation rate in [potentially non-functional] spring alleles (Fig 5.1A). The predominance of a single winter allele indicates that it has been selected for in breeding populations. The presence of this allele in 7 out of 14 spring lines suggest its effect can be overridden by other factors. It is perhaps necessary, but not solely responsible, for a late flowering phenotype. No polymorphisms were identified in *BnFLCA10* likely to impact *COOLAIR* structure or processing (Fig. 5.1B). Two polymorphisms were identified that could affect protein function: an A to C SNP in exon 1 (+58 nt; box 1) which causes an amino acid change from polar Threonine to non-polar Proline, and a G to A SNP at the 5’ splice site in intron 6 (+3771 nt; labelled 3). The latter has previously been reported to cause alternate splicing of *FLC* mRNA and variation in flowering time (Yuan *et al.*, 2009) and was also observed in 4/13 *B. rapa* lines (Fig. S8). The T to C SNP in the second of two highly conserved RY elements in intron 1 (+487 nt) changes the motif from 5’-TGCATG-3’ to 5’-TGTATG-3’ in the ‘winter’ allele. The same mutation, but in the first RY element, is responsible for delayed vernalization in the *A. thaliana vrn8* mutant (Qüesta *et al.*, 2017).

Greater variation exists at *BnFLCA3a* in winter than in spring types, with 12 out of 13 spring cultivars containing the same allele (Fig. 5.2A). This could confer an early

flowering phenotype. There are no polymorphisms exclusive to this allele, but combinations of polymorphisms may be important, as demonstrated for Lov-1 (Coustham et al., 2012). Several polymorphisms were identified that could affect *COOLAIR* processing or splicing (Fig. 5.2B). These include three SNPs in the proximal termination region (labelled 1) which could disrupt addition of the Poly(A) tail. Other SNPs were found in the *COOLAIR* small intron (labelled 2) and *COOLAIR* exon 1 and promoter region (labelled 3), which may influence *COOLAIR* transcription initiation and/or canonical splicing of the small intron (important for FLD-mediated sense downregulation, Marquardt *et al.*, 2014). One of these SNPs causes a change from 5'-CA-3' to 5'-CT-3' (A to T at position +4412); the latter could act as an alternative upstream 3' intron splice site for the proximal *COOLAIR* transcript.

There was high variation within *BnFLCC2* in both spring and winter lines, although some nominal 'alleles' differ by only one base (Fig. 5.3A). Some were found more commonly in winter types and some more commonly in spring types; the polymorphisms that these share may be worth investigating. For example, there are several upstream SNPs that associate with spring rather than winter alleles (-105 nt to -169 nt). Other polymorphisms were identified that may alter *COOLAIR* processing; for example, there is a TATT insertion adjacent to one of the distal terminal exon splice sites (labelled 1; Fig. 5.3B). This insertion was present in a single allele found predominantly (but not exclusively) in winter types. A 5'-CC-3' to 5'-CT-3' mutation (C to T at position +3373; labelled 2) found in both spring and winter cultivars could provide an additional 3' splice site for the proximal small intron. In two winter and four spring cultivars, *FLCC2* could not be amplified from any of the four target regions, indicating that it has been lost in those lines. Presence-absence variation in flowering time genes is thought to be a key driver of phenotypic change (Schiessl *et al*., 2014).

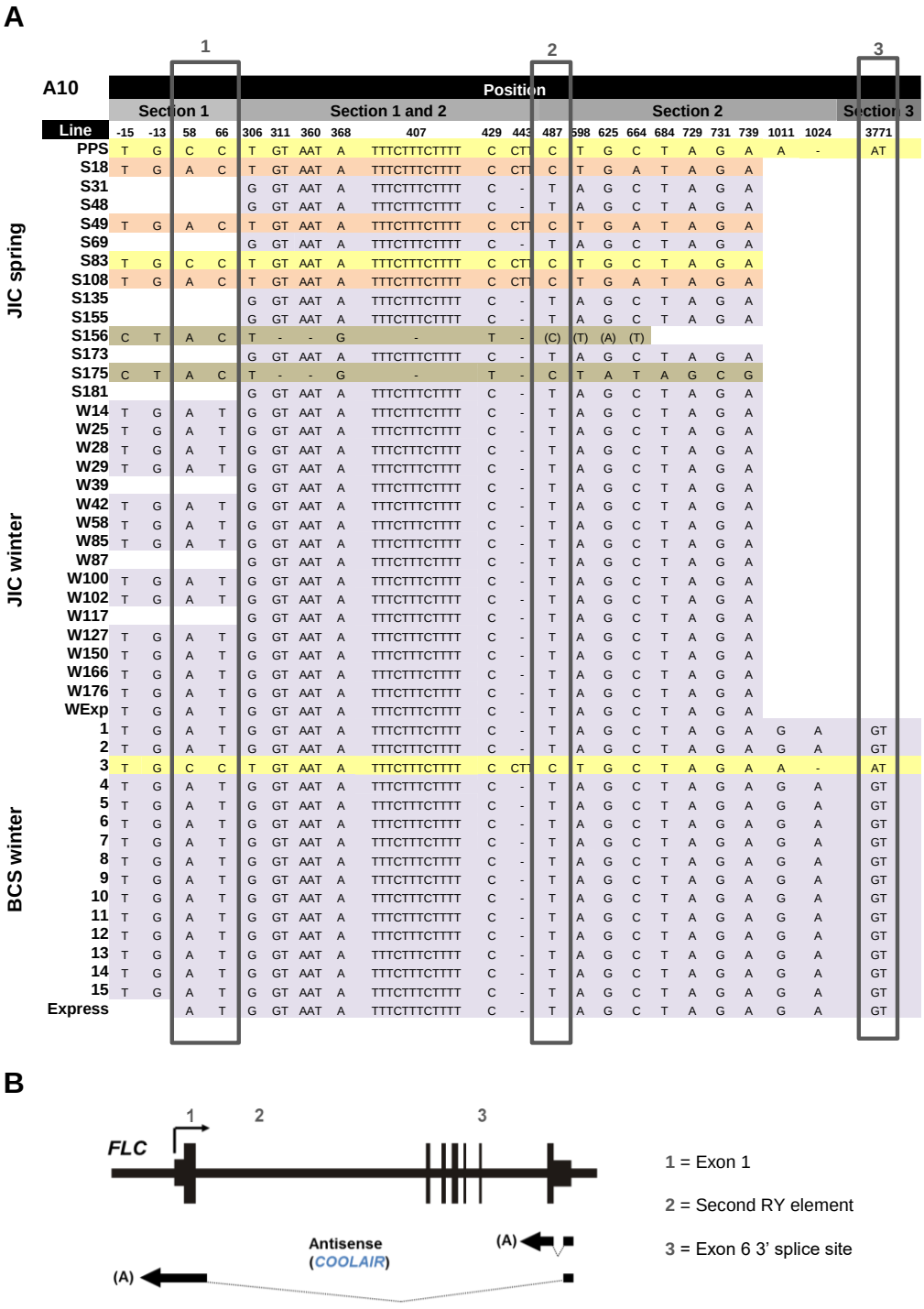


Figure 5.1: Polymorphism mapping in *COOLAIR*-specific regions of *FLCA10* in *B. napus*.
32 winter and 13 spring *B. napus* lines were selected from the JIC diversity or Bayer CropScience (BCS) breeding sets and key sections (1-3) of *FLCA10* sequenced (see Fig. 2.1 for more details). (A) Polymorphisms are positioned with respect to the ATG start site of PPS (+1). PPS refers to the BCS PPS20144 spring line and Express to the BCS Express winter line; sequence data from these were made available by BCS. WExp is the Express line from the JIC diversity set. Candidate *FLC* alleles are grouped by colour; boxed regions highlight polymorphisms of interest. White spaces indicate incomplete or poor sequencing reads. (B) Schematic to show the approximate positions of boxed polymorphisms of interest at the *FLC* gene.

A

	A3a	Position																
		Section 1								Section 3								
	Line	-151	-109	-89	439	461	3898	3930	3967	4131	4136	4143	4410	4412	4530	4547	4551	4635
JIC spring	PPS	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S18	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S31	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	S48	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S49	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S69	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S83	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S108		A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S155	-	A	G	A	-												
	S156	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S173	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S175	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	S181	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
JIC winter	W14	A	?	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	W25	?	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	W28	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	W29	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	W39	A	T	A	A	TTC												
	W42	A	T	A	A	TCC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	W58	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	W85	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	W87	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	W100	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	W102	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	W117	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	W127	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	W150	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	W166	-	A	G	A	-	-	-	T	C	G	T	G	A	T	T	T	A
	W176	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	WExp	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	1	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	2	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	3	-	A	G	A	-												
BCS winter	4	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	5	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	6	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	7	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	8	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	9	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	10				A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	11				A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	12	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	13																	
	14	A	T	A	A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T	C	T	C	G
	15	A	A	G	G	TTC	-	-	T	C	T	T	G	A	T	-	T	A
	Express				A	TTC	CATATACAC	TAGAATAA	A	T	G	A	T	T				

B

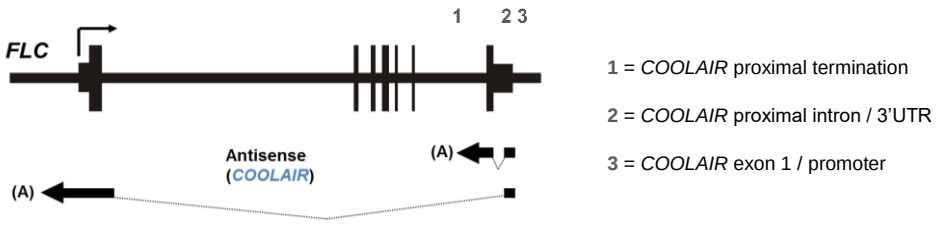


Figure 5.2: Polymorphism mapping in *COOLAIR*-specific regions of *FLCA3a* in *B. napus*.
32 winter and 13 spring *B. napus* lines were selected from the JIC diversity or Bayer CropScience (BCS) breeding sets and key sections (1-3) of *FLCA3a* sequenced (see Fig. 2.1 for more details). (A) Polymorphisms are positioned with respect to the ATG start site of PPS (+1). PPS refers to the BCS PPS20144 spring line and Express to the BCS Express winter line; sequence data from these were made available by BCS. WExp is the Express line from the JIC diversity set. Candidate *FLC* alleles are grouped by colour; boxed regions highlight polymorphisms of interest. White spaces indicate incomplete or poor sequencing reads. (B) Schematic to show the approximate positions of boxed polymorphisms of interest at the *FLC* gene.

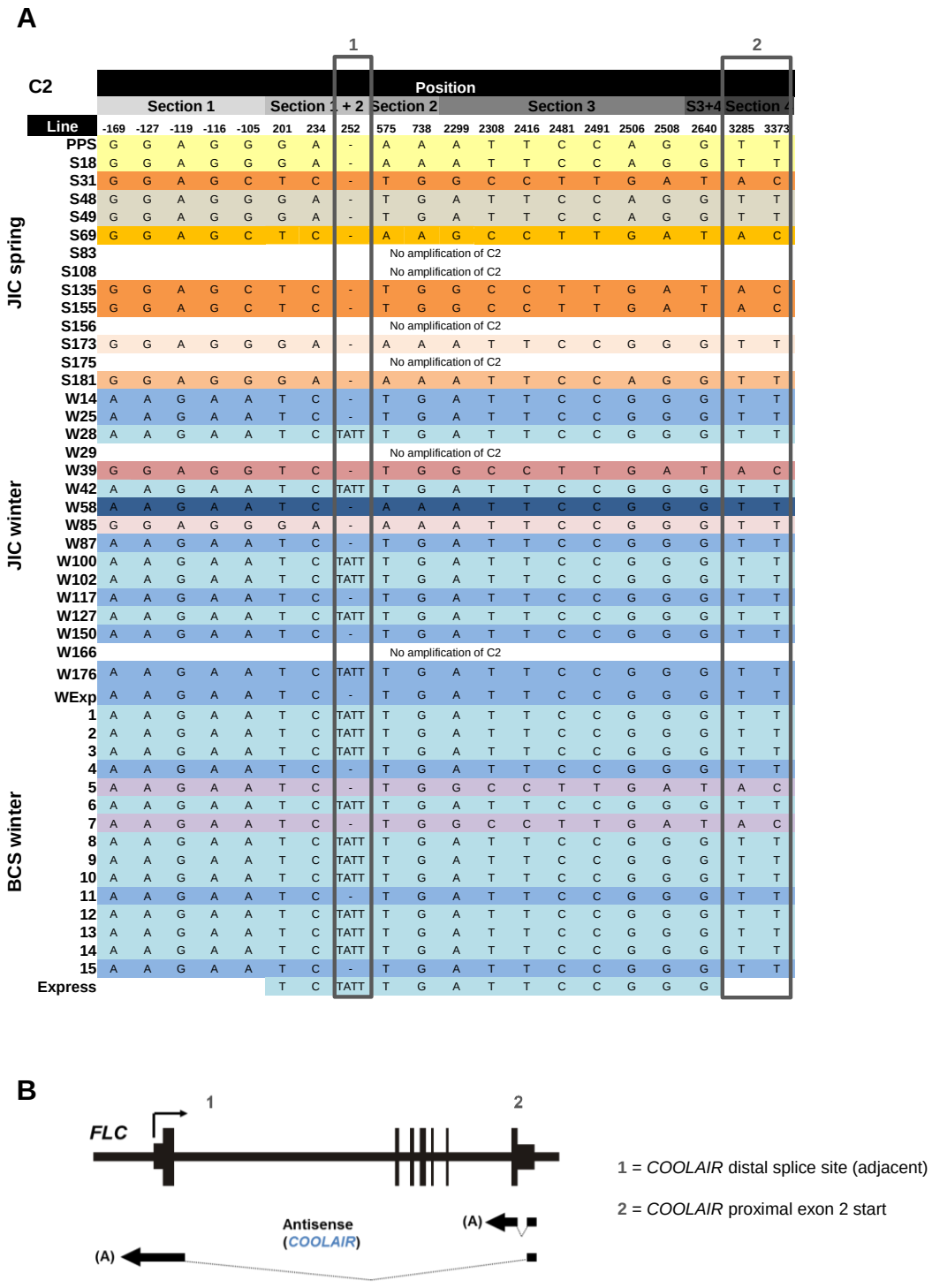


Figure 5.3: Polymorphism mapping in *COOLAIR*-specific regions of *FLCC2* in *B. napus*.
32 winter and 13 spring *B. napus* lines were selected from the JIC diversity or Bayer CropScience (BCS) breeding sets and key sections (1-4) of *FLCC2* sequenced (see Fig. 2.1 for more details). (A) Polymorphisms are positioned with respect to the ATG start site of PPS (+1). PPS refers to the BCS PPS20144 spring line and Express to the BCS Express winter line; sequence data from these were made available by BCS. WExp is the Express line from the JIC diversity set. Candidate *FLC* alleles are grouped by colour; boxed regions highlight polymorphisms of interest. White spaces indicate incomplete or poor sequencing reads; there was no amplification of *FLCC2* in six lines. (B) Schematic to show the approximate positions of boxed polymorphisms of interest at the *FLC* gene.

Polymorphisms with the potential to affect *COOLAIR* processing were investigated further. Eight *Brassica napus* cultivars from the JIC core diversity set were selected for qualitative analysis of *FLCA3a* and *FLCC2 COOLAIR* transcripts by RT-PCR: four winter (W14, W42, W58, W150) and four spring (S48, S69, S135, S173). These cultivars were chosen because they contained alleles with polymorphisms of interest. Total RNA from non-vernalized leaf tissue were equalised for the eight lines using the *UBC* reference gene (Fig. 5.4A), and *BnFLCA3a* (Fig. 5.4B) and *BnFLCC2* (Fig. 5.5) proximal and distal *COOLAIR* transcripts amplified by touch-down nested reverse transcription (RT)-PCR.

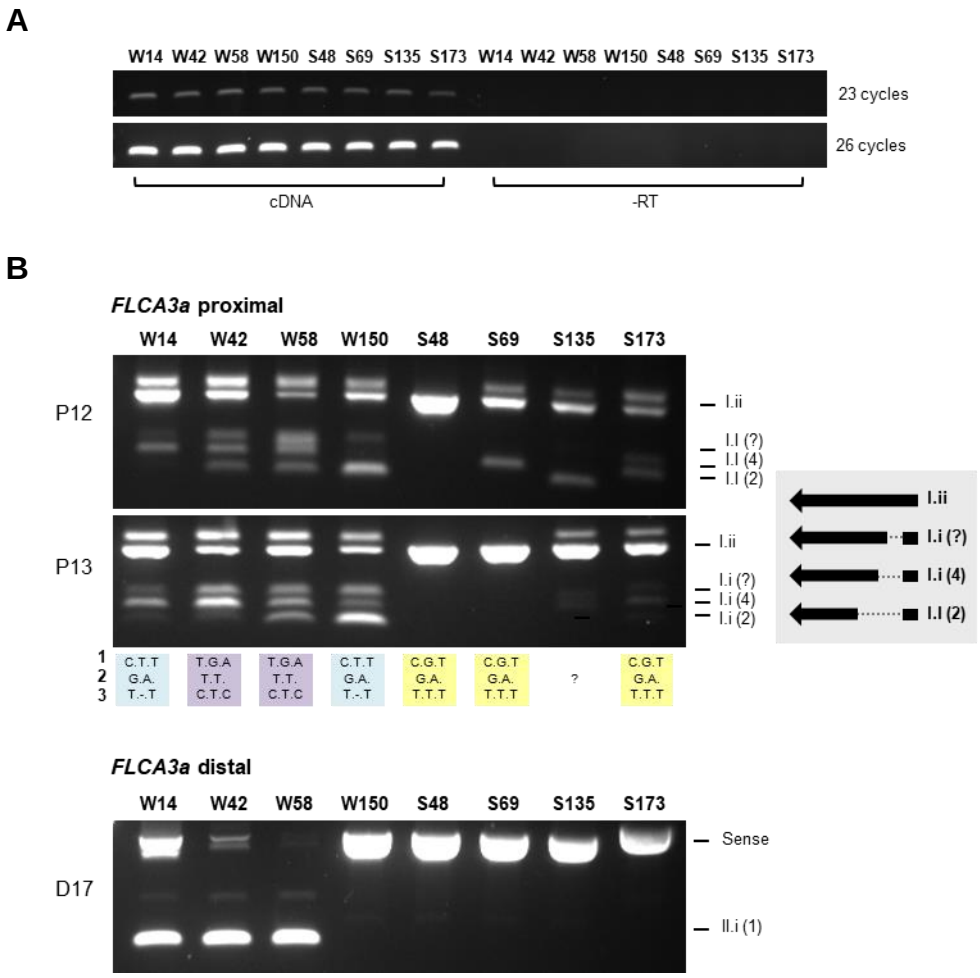


Figure 5.4: *FLCA3a COOLAIR* splicing in spring and winter *B. napus* lines. Touch-down nested RT-PCRs of *BnFLCA3a COOLAIR* proximal and distal transcripts from the leaf tissue of eight non-vernalized *B. napus* cultivars. W14, W42, W58 and W150 are winter lines, whereas S48, S69, S135 and S173 are spring lines. (A) Amplicons were first equalised for the *UBC* reference gene at different cycle numbers (23 and 26 cycles shown), where the -RT columns are loaded with the no reverse transcriptase (DNA contamination) control. (B) Gel images for the second nested RT-PCR are shown, with proximal specific primers P12 and P13 (top), and distal-specific primers D17 (bottom). Coloured boxes indicate the *FLC* allele and polymorphisms of interest found in that cultivar, with 1-3 referring to the same in Fig. 5.2. Spliceoforms were categorised by size according to previously identified *A. thaliana* and *Brassica* classes (Fig. 3.1 and Fig. 4.7), although one form (I.i(?)) was not recognised and may be new. A schematic of predicted proximal splicing patterns is given to the right. The two upper proximal bands appeared to have identical nucleotide sequences (pertaining to proximal I.ii) after cloning and sequencing; it may be that only one was successfully cloned, or they may have different structures and consequently different migration speeds on the gel. 'Sense' indicates where primers amplified from the *FLCA3a* sense transcript.

Three regions of SNPs were identified at the 3' end of *BnFLCA3a*, with the potential to affect proximal *COOLAIR* processing (Fig. 5.2). Three of the four analysed summer cultivars contained the same *BnFLCA3a* allele (with S135 being unknown), whereas the winter lines contained one of two different *BnFLCA3a* alleles. Three spliced proximal isoforms were expressed in the winter lines, in addition to the unspliced Class I.ii transcript; these were more difficult to detect in the spring lines, but all three were expressed (at a low level) in S173 (Fig. 5.4). Distinct splicing patterns were not detected between lines W42/58 and W14/150, indicating that the polymorphisms between their alleles in regions 1-3 do not affect *COOLAIR* processing. The A to T SNP at +4412 nt in W42 and W58 did not introduce a new proximal receptor splice site in the small intron. The *BnFLCA3a* distal transcript was only detected in winter cultivars (except W150); this was previously observed when comparing winter Major and Express with spring Stellar and Westar (Fig. S7). Given the interconnection between sense and antisense, repression of the *BnFLCA3a* locus in spring cultivars could explain low *COOLAIR* proximal and distal expression. This is supported by a lack of distal-specific polymorphisms that could influence *COOLAIR* processing. Indeed, *BnFLCA3a* expression levels are lower in S48, S69, S135, and S173 than in W14, W42, W58, and W150 (Fig. 4.2).

Polymorphisms were identified at the 5' and 3' end of *BnFLCC2* with the potential to affect distal and proximal *COOLAIR* processing, respectively (Fig. 5.3). There is a TATT addition adjacent to the distal terminal exon splice site in W42, which correlates with unsuccessful amplification of the distal transcript (Fig. 5.5B), despite similar *BnFLCC2* sense expression levels in W14 (without the TATT addition) and W42 (with the TATT addition) (Fig. 4.2). In four other lines, however, the distal was only amplified from 1 of 2 primer sets (not both) signifying suboptimal experimental conditions. Other *B. napus* cultivars with the TATT polymorphism should be investigated to determine whether this is a trend. The C to T SNP at +3373 nt, within the proximal small intron, has the potential to introduce an early proximal exon 2 start site in S48, S173, and the four winter cultivars (Fig. 5.3). Sequencing of RT-PCR products confirmed that this was true, with three splicing isoforms present in lines with the C to T SNP, and only two isoforms (one dominant) in lines without (Fig. 5.5A). All spring lines in this experiment have roughly similar *BnFLCC2* expression levels (Fig. 4.2), precluding this as the causative factor.

In addition to splicing, naturally-occurring *cis* polymorphisms also have the potential to alter secondary structure. We previously described a G to C SNP in the *B. rapa* PTM cultivar which could disrupt base pairing in helix H4, potentially shortening or

destabilising this helix compared with the SHAPE-probed R018 structure (Fig. 4.9D; Chapter 4). The difference is small, but could be partially responsible for differences in *FLCA3a* sense expression and consequently flowering time.

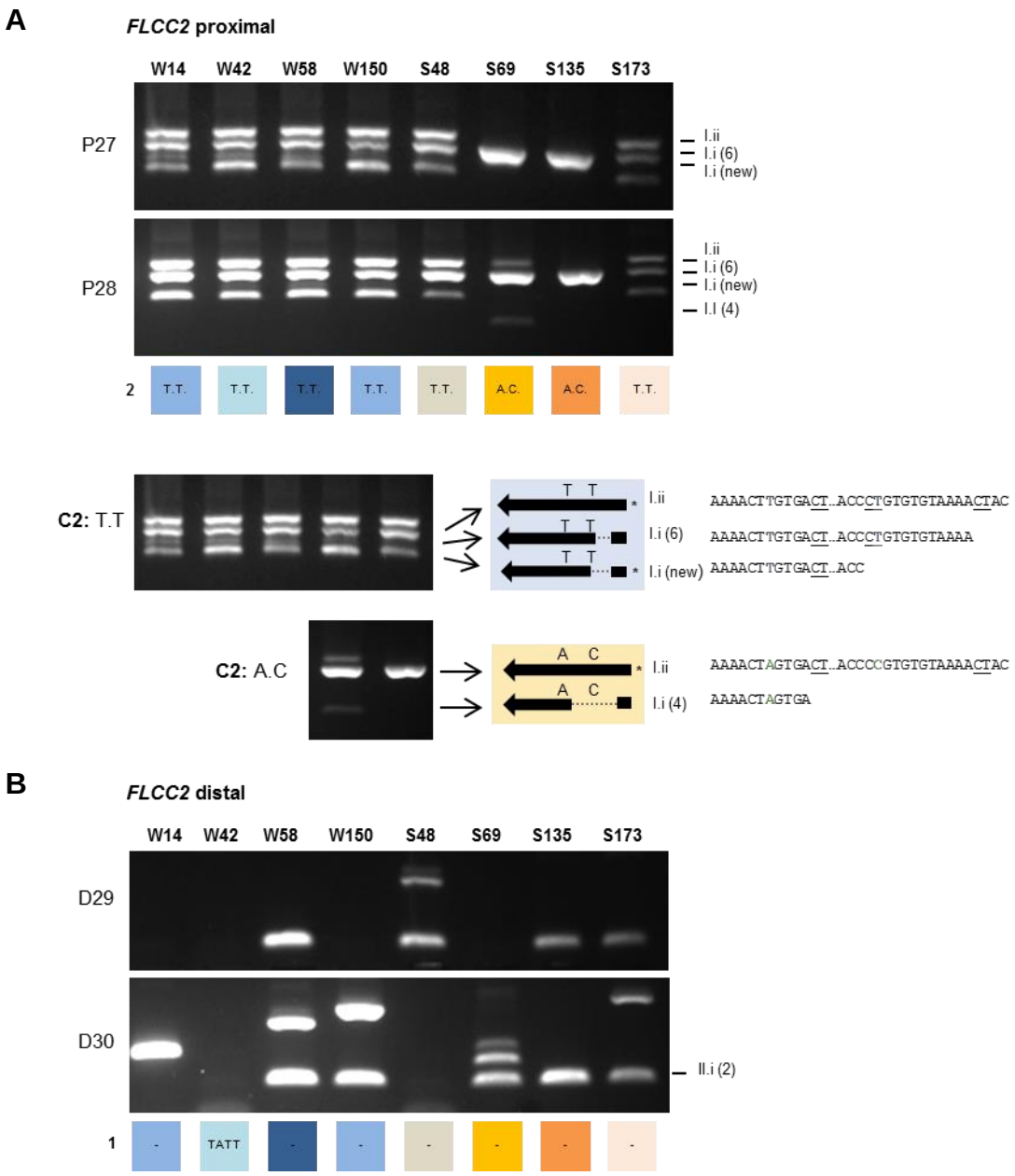


Figure 5.5: *FLCC2 COOLAIR* splicing in spring and winter *B. napus* lines. Touch-down nested RT-PCRs of *BnFLCC2 COOLAIR* proximal and distal transcripts from the leaf tissue of eight non-vernalized *B. napus* cultivars. W14, W42, W58 and W150 are winter lines, whereas S48, S69, S135 and S173 are spring lines. Amplicons were first equalised for the *UBC* reference gene at different cycle numbers (Fig. 5.4A). Gel images for the second nested RT-PCR are shown, with proximal-specific primers P27 and P28 (A), and distal-specific primers D29 and D30 (B). Coloured boxes indicate the *FLC* allele and polymorphisms of interest found in that cultivar, with 1 and 2 referring to the same in Fig. 5.3. Proximal spliceoforms were sequenced and categorised according to *A. thaliana* classes and *Brassica* classes (Fig. 3.1 and Fig. 4.7). The distal II.i (2) spliceoform was categorised by size, as above. Schematics of proximal splicing patterns are given beneath the main proximal gel image; * indicates the predominant isoform(s).

5.2.2 Designing FLCA3a *and* FLCA10 *constructs with altered distal splicing patterns*

Use of alternate distal 3' acceptor sites has been linked to differences in *COOLAIR* regulation of *AtFLC* in *cis* (Li *et al.*, 2015). *BrFLCA3a* uses the canonical Col-0 distal splice site, whereas *BrFLCA10* uses the Var2-6 splice site. Use of the latter is responsible for altered distal splicing, high *AtFLC* expression and late flowering in the *A. thaliana* Var2-6 accession (Li *et al.*, 2015). We hypothesise that use of this splice site drives higher levels of *BrFLCA10* transcription (compared with *BrFLCA3a*) within a single *Brassica rapa* plant. To test this, we set up experiments designed to reciprocally switch distal splice sites in *BrFLCA10* and *BrFLCA3a*. *Brassica FLC* constructs were transformed into *A. thaliana* Col *FRI flc-2* (without a functional *FLC*) to analyse their effect without the presence of other (potentially compensatory) *Brassica FLC* homologues.

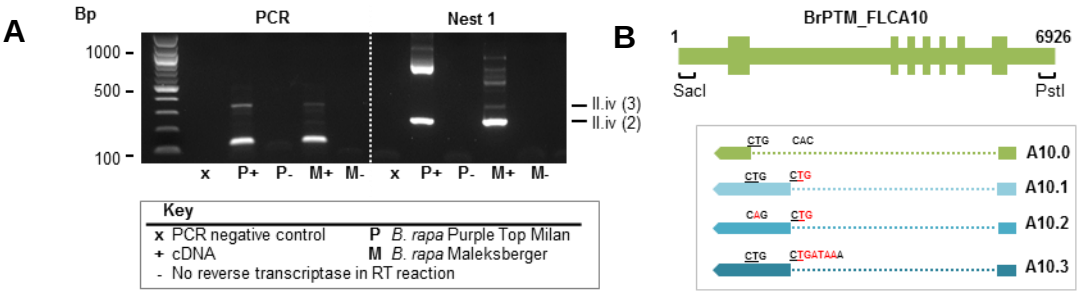
Previously, *COOLAIR* expression was detected from *FLCA10* in R018, a spring *B. rapa* cultivar. Sequencing of *BrFLCA10* in R018 revealed that it contained the G to A SNP at the 5' splice site in intron 6, responsible for defective FLC proteins (Yuan *et al*., 2009). This would negate differences in flowering time from *COOLAIR*. This region was sequenced in 13 additional *B. rapa* cultivars (Fig. S8). From these, two without the G to A SNP, the rapid-cycler Maleksberger and winter Purple Top Milan (PTM), were selected for full sequencing of *BrFLCA10*. These contained highly similar *BrFLCA10* alleles [there were a low number of polymorphisms between the two] without any SNPs known to affect the full-length FLC protein. *COOLAIR* proximal (not shown) and distal splicing were the same as in R018, with the Class II.iv (2) form detected in both cultivars (Fig. 5.6A). The Class II.iv (3) form with an extra exon was also detected in Maleksberger. The *BrFLCA10* allele from PTM was selected for subsequent experiments. This was because it only (or predominantly) expressed the Class II.iv (2) form; thus negating structural complementation (if it exists) from the extra exon.

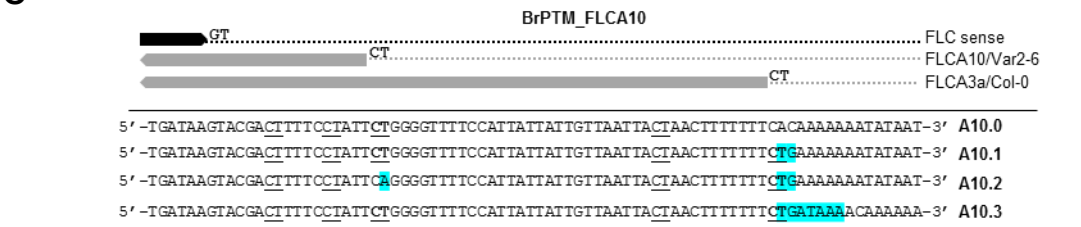
The full-length *BrFLCA10* wildtype allele (designated A10.0) was synthesised from 1.5 kb upstream of the translation start site to 1.1 kb downstream of the translation stop site. Three variants of the *BrFLCA10* wildtype allele, A10.1, A10. 2 and A10.3, were designed to shift distal intron splicing upstream from the wildtype Var2-6 to the Col-0 site (Fig. 5.6B). Pair-wise sequence alignment with *AtFLC* allowed us to predict where this should be. Construct A10.1 reinstates the Col-0 splice site by mutating 3'-CAC-5' to 3'-CTG-5'; where 3'-CT-5' is a canonical antisense splice site (Fig. 5.6C). The adjacent C to G mutation is important because a change from G to T in this position in Var2-6 induces use of the downstream splice site (Li *et al*., 2015). Sequence context can be critical for splicing

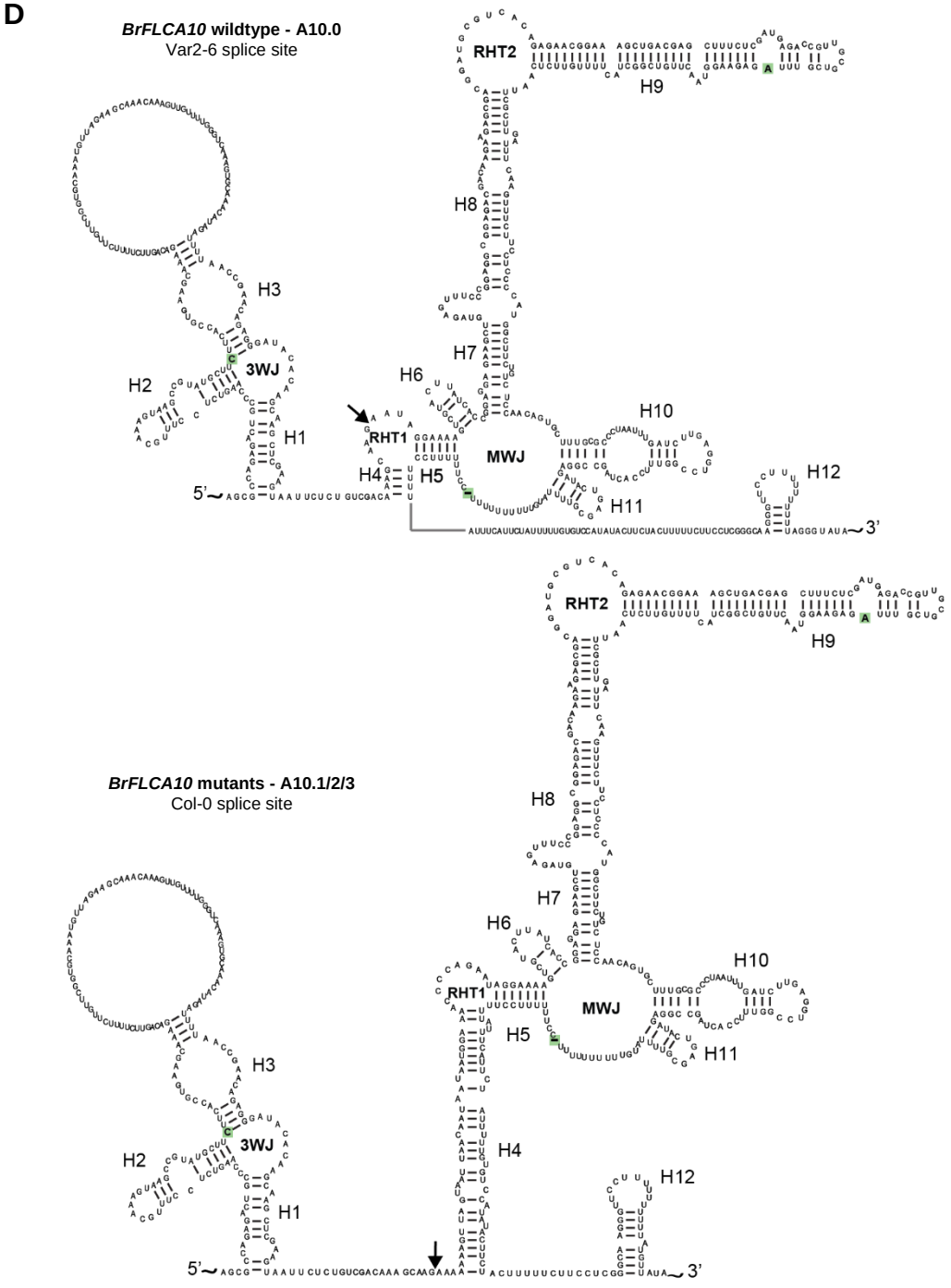
therefore A10.3 reinstates the 3'-CTG-5' antisense splice site, plus surrounding sequence from *BrFLCA3a* (Fig. 5.6C). The Var2-6 splice site is left intact in constructs A10.1 and A10.3 because the Col-0 site is considered stronger; it is preferentially used where present in *Arabidopsis* accessions and *Brassica* species. The A10.2 construct has a mutated Var2-6 site in addition to the reinstated Col-0 splice site, in case the above does not hold true within the *BrFLCA10* sequence context.

COOLAIR distal secondary structure was predicted for the expected spliced forms of the *BrFLCA10* wildtype and mutant constructs (Fig. 5.6D). The secondary structure for the wildtype A10.0 construct (Fig. 5.6D, top) is identical to that predicted for R018 in Fig. 4.8. There are only three polymorphisms within this region between PTM and R018; these are highlighted in green and do not disrupt secondary structure (Fig. 5.6D). This structure contains an extremely short (4 bp) helix H4, which has been associated with weak negative regulation of the sense. When the Col-0 splice site is used (Fig. 5.6D, bottom), a long (25 bp) helix H4 is reinstated. Although in length this is closer to the *Arabidopsis* Var2-6 than Col-0 secondary structures, it is longer than the canonical *BrFLCA10* H4 structure. It more closely resembles *BrFLCA2* H4 than *BrFLCA3a*, which has an 11 bp deletion in this region, and therefore may be a strong regulator of the sense. The only other difference between the two structures is that the length of the linker region between the 5' and 3' structural domains is longer in the mutants, 25 nt versus 14 nt.

Figure 5.6: *Brassica FLCA10* construct design (next page). (A) Touchdown nested RT-PCRs to detect distal *COOLAIR* transcripts from *FLCA10* in *B. rapa* Purple Top Milan (PTM) and Maleksberger cultivars (non-vernalized leaf tissue). Initial and first nested PCR are shown. Spliceoforms were sequenced and categorised according to *A. thaliana* and *Brassica* classes (Fig. 3.1 and Fig. 4.7); *FLCA10* from PTM was selected for constructs. (B) Schematic of the *B. rapa* PTM *FLCA10* constructs. A10.0 represents the wildtype construct and spliceoform; A10.1, A10.2, and A10.3 represent the mutated constructs (splice site mutations in red) and intended spliceoforms, which uses the Col-0 distal splice site. Potential antisense splice sites are underlined. (C) Nucleotide sequence of the distal *COOLAIR* splice site region in the four constructs. Potential antisense splice sites are underlined, canonical *A. thaliana* sites are in bold, and mutations highlighted in blue. (D) Predicted secondary structures of the wildtype A10.0 distal *COOLAIR* (top) which uses the Var2-6 splice site, and of the A10.1/2/3 mutants (bottom) if they use the Col-0 splice site, as intended. Arrows mark exon-exon boundaries. These structures are predicted from *B. rapa* PTM nucleotide sequence, rather than the R018 sequence used in Fig. 4.9; there are three polymorphisms (highlighted in green) at *COOLAIR* between the two, but these are unlikely to alter structure. Grey lines are there to facilitate drawing of the 2D structure.


A
Bp
PCR
Nest 1
1000
500
100
x P+ P- M+ M- x P+ P- M+ M-
II.iv (3)
II.iv (2)
Key
x PCR negative control
+ cDNA
- No reverse transcriptase in RT reaction
P B. rapa Purple Top Milan
M B. rapa Maleksberger
B
BrPTM_FLCA10
1
6926
SacI
PstI
CTG CAC
CTG CTG
CAG CTG
CTG CTGATAAA
A10.0
A10.1
A10.2
A10.3


C
BrPTM_FLCA10
GT
CT
CT
FLC sense
FLCA10/Var2-6
FLCA3a/Col-0
5′-TGATAAGTACGACTTTTCCTATTCTGGGGGTTTTCCATTATTATTGTTAATTACTAACTTTTTTTCACAAAAAAATATAAT-3′ A10.0
5′-TGATAAGTACGACTTTTCCTATTCTGGGGGTTTTCCATTATTATTGTTAATTACTAACTTTTTTTCTGAAAAAAATATAAT-3′ A10.1
5′-TGATAAGTACGACTTTTCCTATTCAGGGGGTTTTCCATTATTATTGTTAATTACTAACTTTTTTTCTGAAAAAAATATAAT-3′ A10.2
5′-TGATAAGTACGACTTTTCCTATTCTGGGGGTTTTCCATTATTATTGTTAATTACTAACTTTTTTTCTGATAAACAAAAAA-3′ A10.3


D
BrFLCA10 wildtype - A10.0
Var2-6 splice site
RHT2
H9
H8
H3
H7
H6
H10
3WJ
H2
H1
RHT1
MWJ
H4
H5
H11
H12
5′
3′
BrFLCA10 mutants - A10.1/2/3
Col-0 splice site
RHT2
H9
H8
H7
H6
H10
RHT1
MWJ
H5
H11
H3
3WJ
H2
H4
H12
H1
5′
3′

The reverse was carried out for *BrFLCA3a*, with constructs designed to switch distal splicing from the Col-0 to the Var2-6 site. *BrFLCA3a* alleles were sequenced in *B. rapa* PTM and Maleksberger, as for *FLCA10*. *COOLAIR* distal II.i (1) transcripts, previously observed in R018, were detected under non-vernalized conditions in both cultivars (Fig. 5.7A).

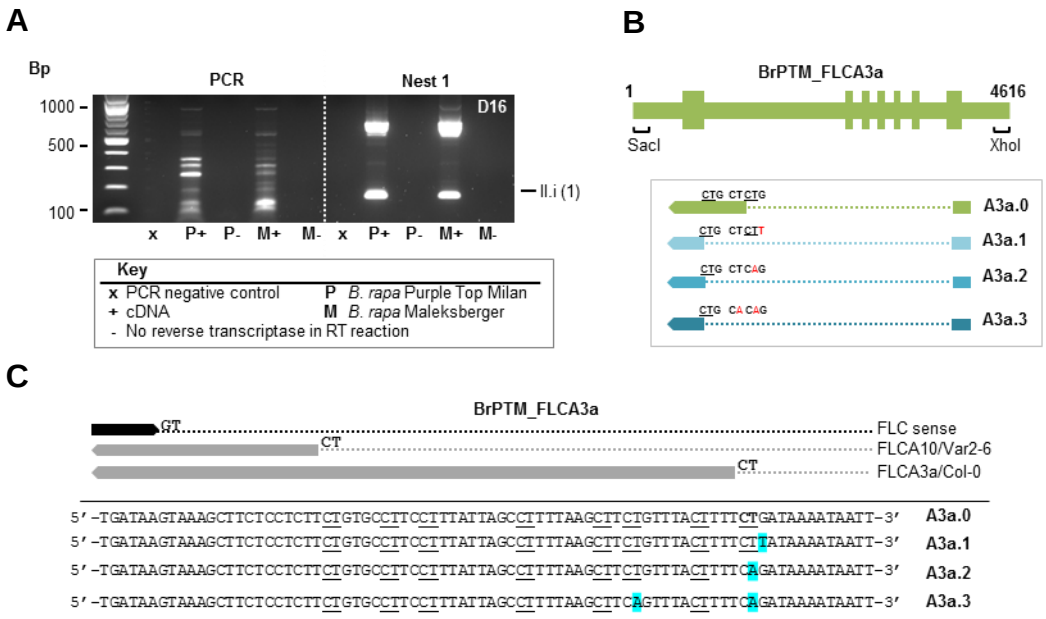


Figure 5.7: *Brassica FLCA3a* construct design. (A) Touchdown nested RT-PCRs to detect distal *COOLAIR* transcripts from *FLCA3a* in *B. rapa* Purple Top Milan (PTM) and Maleksberger cultivars (non-vernalized leaf tissue), using primer set 16. Initial and first nested PCR are shown. Spliceoforms were sequenced and categorised according to *A. thaliana* and *Brassica* classes (Fig. 3.1 and Fig. 4.7); *FLCA3a* from PTM was selected for constructs. (B) Schematic of the *B. rapa* PTM *FLCA3a* constructs. A3a.0 represents the wildtype construct and spliceform; A3a.1, A3a.2, A3a.3 represent the mutated constructs (splice site mutations in red) and intended spliceforms, which uses the Var2-6 distal splice site. Potential antisense splice sites are underlined. (C) Nucleotide sequence of the distal *COOLAIR* splice site region in the four constructs. Potential antisense splice sites are underlined, canonical *A. thaliana* sites are in bold, and mutations highlighted in blue.

The *BrFLCA3a* allele from PTM was selected for the following experiments, allowing comparison between *BrFLCA10* and *BrFLCA3a* homologues from the same cultivar. PTM was heterozygous for *BrFLCA3a* but both alleles were expected to produce functional FLC proteins. The full-length selected *BrFLCA3a* wildtype allele (designated A3a.0) was synthesised from 0.97 kb upstream of the translation start site to 0.71 kb downstream of the translation stop site. Further up- or down-stream regions were not included in the constructs because they encroached on neighbouring genes. Three variants, A3a.1, A3a.2 and A3a.3, were designed to shift distal intron splicing downstream from the wildtype Col-0 to the Var2-6 splice site (Fig. 5.7B). The Var2-6 splice site has been conserved in *BrFLCA3a*, making these experiments theoretically easier than the *BrFLCA10* experiments

because a new splice site does not need to be synthesised. Constructs A3a.1 and A3a.2 are designed to inhibit use of the Col-0 splice site. A3a.1 has a mutation adjacent to the splice site, converting 3'-CTG-5' to 3'-CTT-5'; this mirrors the *A. thaliana* Var2-6 mutation (Fig. 5.7C). A3a.2 mutates the actual splice site, converting 3'-CTG-5' to 3'-CAG-5', in case the A3a.1 mutation does not work (Fig. 5.7C). A3a.3 additionally mutates another downstream 5'-CTG-3' splice site, in case this second 'strong' splice site is used over the Var2-6 site (Fig. 5.7C). There are other potential antisense splice sites in this region (underlined in Fig. 5.7C) and therefore it is possible that the designed constructs will not use the intended Var2-6 site.

COOLAIR distal (Class II.i) secondary structures were predicted for the expected spliceoforms of the *BrFLCA3a* wildtype and mutant constructs (Fig. 5.8). The chemically-probed secondary structure is given for *BrFLCA3a* from R018 (top) and the predicted structure for PTM (middle). SNPs between the two are highlighted in green in the PTM structures. The G to C SNP in helix H4, described above and in Chapter 4, has the potential to disrupt secondary structure, shortening H4 and shifting the angle of the first r-turn. This is hypothesised to give rise to a weaker distal *COOLAIR*. The proposed shift to the Var2-6 splice site (bottom) would cause even greater disruption to secondary structure, with loss of H4, the first r-turn and the multi-way junction. The A3a.1, A3a.2 and A3a.3 mutant constructs were designed to produce this [presumably] non-functional *COOLAIR*.

BrFLCA10 and *BrFLCA3a* constructs were synthesised in the pL0M vector, and cloned into pSLJ755I6 for transformation into Col *FRI flc-2* plants via *Agrobacterium*-mediated floral dip (Table 5.1; Fig. 2.2). Col *FRI flc-2* is a transgenic *Arabidopsis* line that contains *FRIGIDA* but no functional *FLC* allele. Positive transformants were identified by BASTA selection at T1. Genotyping and copy number analysis were performed for 60+ positive T1 transformants per construct (Fig. 5.9). There was weak positive correlation between *BrFLCA10* and *BrFLCA3a* copy number and flowering time; transformants with higher copy numbers tended to flower later (Fig. 5.9). Correlation may have been confounded by relative activity of genomic loci. Genotyping revealed that a low number of plants which survived BASTA selection did not contain the *BrFLCA10* or *BrFLCA3a* transgene. One possibility is that incomplete transgenes, containing the BASTA-resistance gene but not the *FLC* construct, were transformed at a low level. Copy number detection was specific to the BASTA-resistance gene; therefore some with multiple copies may be false positives with no effect on flowering time. Single copy positive transformants for each construct were carried through to T2 for further analysis.

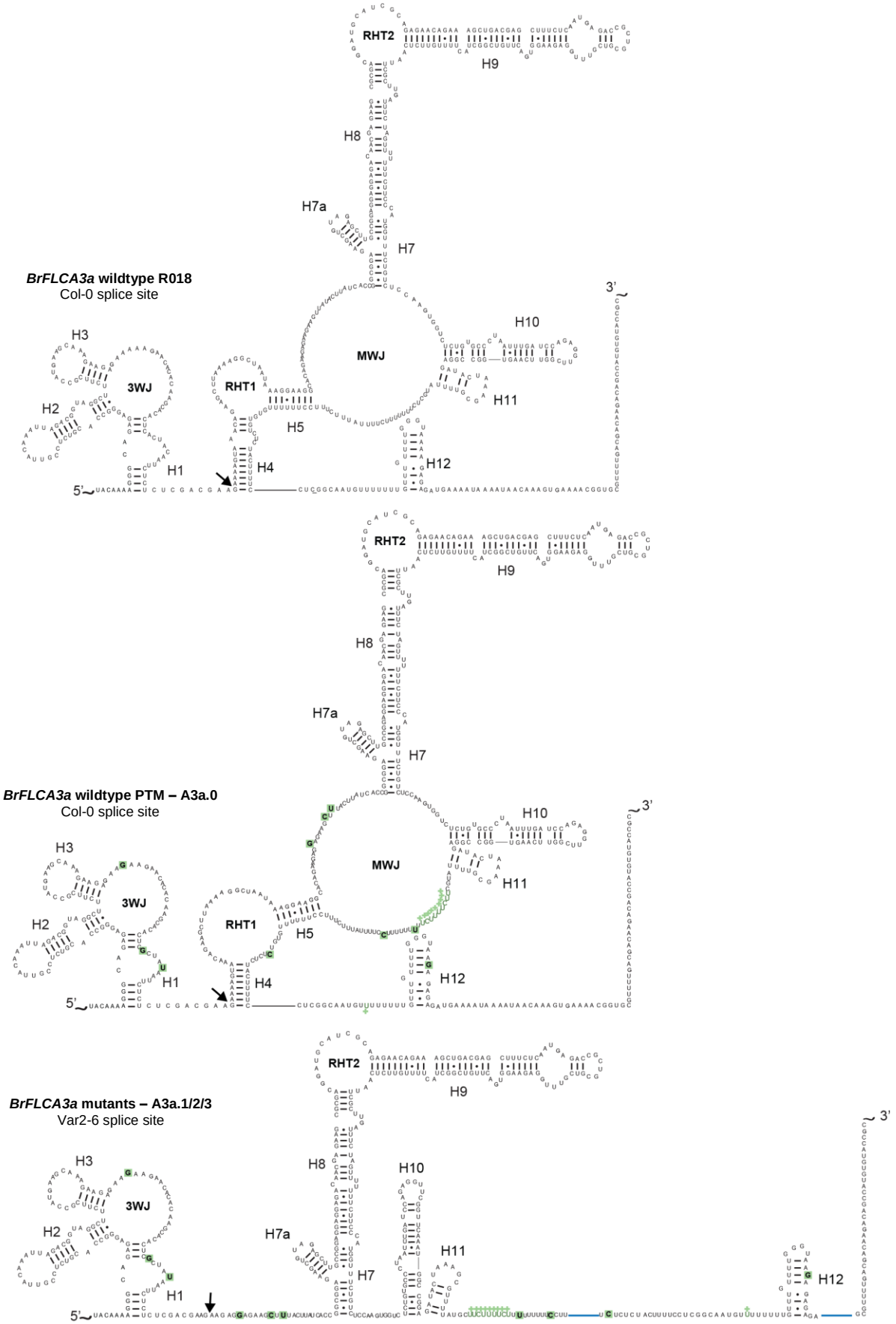


Figure 5.8: Predicted distal *COOLAIR* secondary structure changes in the *BrFLCA3a* mutants. Chemically-probed secondary structure of the distal *COOLAIR* from R018 (top) and predicted structure of the altered form in PTM (middle), both of which use the Col-0 splice site. Nucleotides highlighted in green represent SNPs between the two sequences, + represents insertions, and – deletions. Arrows mark exon-exon boundaries. Below is the predicted structure of the A3a.1/2/3 mutants if they use the intended Var2-6 splice site. Blue lines mark where sections of single-stranded (non-structural) sequence have been removed to save space; grey lines are there to facilitate drawing of the 2D structure. Structure of the *BrFLCA3a* mutant isoform with full-length sequence is given in Fig. S9.

ID	Construct	Vector	Background	T1 Lines	T2 Lines	T2 Plants
A10.0	*B. rapa* PTM FLCA10 WT	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	60	13	10
A10.1	*B. rapa* PTM FLCA10 v1	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	60	21	10
A10.2	*B. rapa* PTM FLCA10 v2	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	60	18	10
A10.3	*B. rapa* PTM FLCA10 v3	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	60	19	10
A3a.0	*B. rapa* PTM FLCA3a WT	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	72	15	10
A3a.1	*B. rapa* PTM FLCA3a v1	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	72	21	10
A3a.2	*B. rapa* PTM FLCA3a v2	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	72	18	10
A3a.3	*B. rapa* PTM FLCA3a v3	PSLJ755I6	*A. thaliana* Col *FRI flc-2*	72	11	10

Table 5.1: *BrFLCA10* and *BrFLCA3a* construct details. Number of lines analysed per construct at T1 and T2 stage is given, and the number of plants analysed per line at T2.

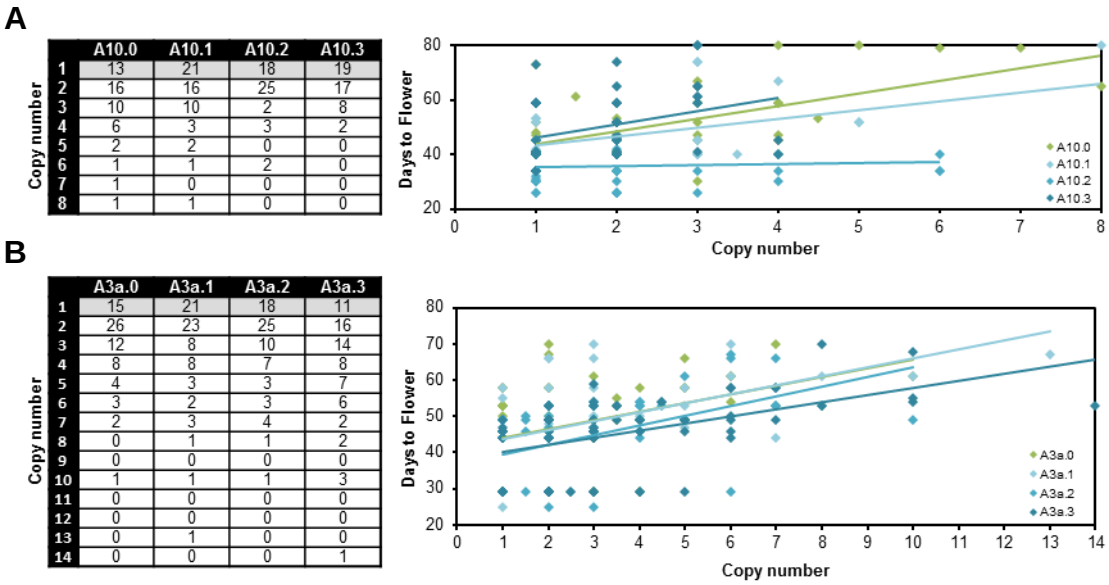


Figure 5.9: Copy number analysis of transgenic lines with *BrFLCA10* (A) and *BrFLCA3a* (B) transgenes. Copy numbers (left) of *BrFLCA10* (A) and *BrFLCA3a* (B) transgenes after *Agrobacterium*-mediated floral transformation in T1 plants, single copy lines were carried forwards to T2. Graphs (right) show relationship between copy number and days to flower, divided by construct type.

5.2.3 The FLCA3a *allele from* B. rapa *PTM might express a non-functional distal* COOLAIR

BrFLCA3a lines were vernalized for 10 days at T1. This was intended to speed up T1 flowering and consequently harvesting times. Some lines flowered before vernalization; these were predicted to be negative transformants (as described above). In fact, genotyping revealed that the majority were positive for the *FLCA3a* transgene. There were no clear flowering time differences between A3a.0 wildtype and A3a.1/2/3 mutant lines with multiple copies (Fig. 5.10A, top). A potential trend towards earlier flowering was observed for single copy lines containing A3a.2 or A3a.3 constructs (Fig. 5.10A, bottom). This was the opposite effect to that expected (late flowering). It is possible that the short 10 day vernalization period overrode flowering time differences at T1.

Non-vernalized and four weeks vernalized flowering times were recorded for T2 plants (Fig. 5.10B). Flowering occurred ~40-50 days after sowing for all lines under non-vernalized conditions, with no obvious differences between constructs. Four weeks vernalization sped up flowering by ~10 days. Vernalized plants flowered at around ~30 days, similar to the Col *FRI flc-2* (no transgene) controls. This suggests vernalization requirement was almost saturated. A few lines were not fully vernalized i.e. A3a.0 29, A3a.1 35 and A3a.2 22. Differences between lines containing the same construct could be related to genome position effects (Matzke and Matzke, 1998).

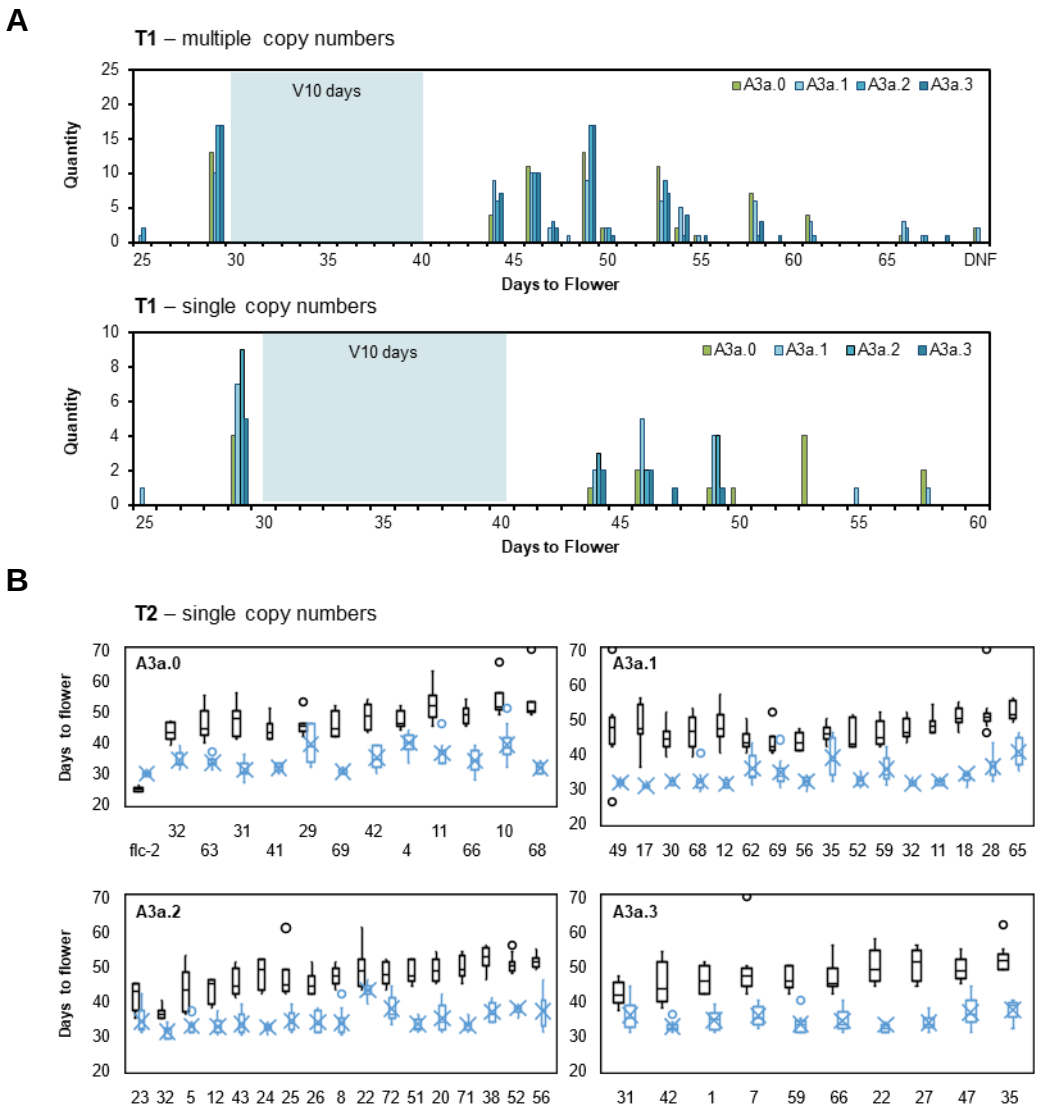


Figure 5.10: Flowering times of *A. thaliana* transgenic plants with *BrFLCA3a* wildtype and mutant transgenes. (A) Flowering times for T1 multi-copy (top) and single-copy (bottom) lines from each construct. Blue box represents a period of 10 days vernalization, although some plants had flowered before this. (B) Non-vernalized (black) and vernalized (blue) flowering times for T2 single-copy lines for each construct, and for the *A. thaliana* Col *FRI flc-2* control (see A3a.0 graph, top-left); 10 plants were analysed for each line. Note that A3a.0 is the wildtype *FLCA3a* construct from *B. rapa* PTM; A3a.1, A3a.2 and A3a.3 are mutant constructs.

As there were no obvious differences in flowering time phenotype, we checked to see whether our polymorphisms had created the intended changes in splicing. *COOLAIR* distal splicing was analysed by the developed method (touchdown nested RT-PCRs) from pooled *BrFLCA3a* lines for each construct (Fig. 5.11A). A ~200 bp nested PCR product represents the canonical distal II.i (1) transcript from the wildtype A3a.0 construct. Similar sized products are detected from the three variant constructs; however, multiple products around this size from A3a.1 and A3a.3 were visible by eye on the gel under UV fluorescence. A larger ~300 bp PCR product was additionally amplified from A3a.0 and A3a.2. Sequencing revealed that the wildtype A3a.0 distal transcript uses the canonical Col-0 splice site as expected, whereas the three mutant constructs do not (Fig. 5.11B). Instead, they each use up to two alternate downstream splice sites. None of the detected transcripts use the Var2-6 splice site (as was intended), despite its conservation. This suggests that sequence context is more important for splice site choice than conservation of a 'strong' site. A3a.1, however, uses a splice site just 4 bp upstream of the Var2-6 site; this transcript therefore has similar architecture to Class II.iv (2). It is interesting that this is used in A3a.1, but not in A3a.2 or A3a.3. In the latter two, the Col-0 splice site was directly mutated; this must somehow affect how nearby splice sites are perceived by the splicing machinery as it scans the gene. Other A3a.1 transcripts and some A3a.2 transcripts use the downstream 3'-CTG-5' antisense splice that was mutated in A3a.3. Mutation of this in A3a.3 promotes use of an upstream splice site (also used by some A3a.2 transcripts) only 3 bp downstream of the Col-0 site. The ~300 bp PCR product detected in A3a.0 and A3a.2 represents a Class II.ii transcript.

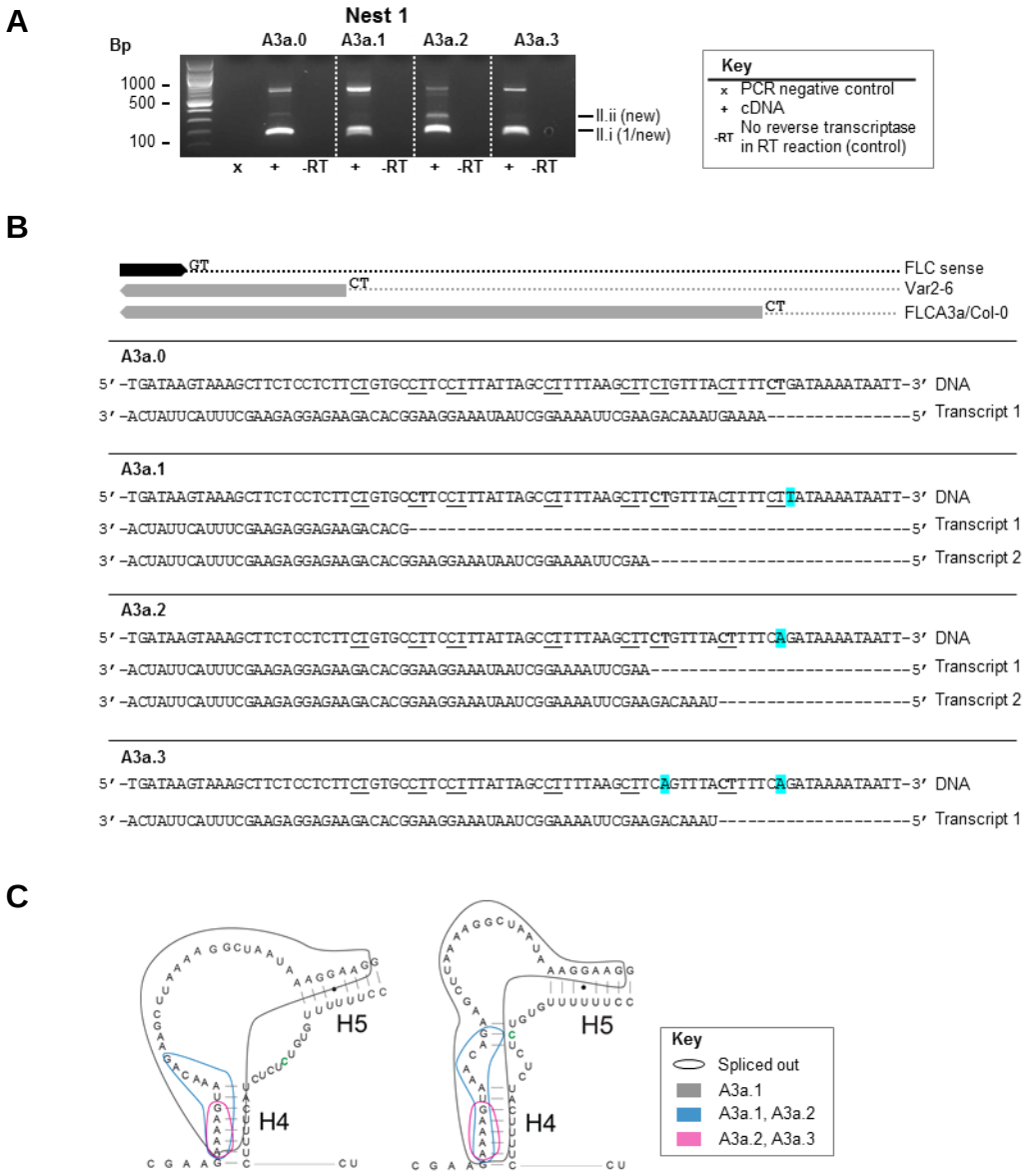


Figure 5.11: Splicing of the distal *COOLAIR* from *BrFLCA3a* transgenes. (A) Touchdown nested RT-PCRs to detect distal *COOLAIR* transcripts from the *FLCA3a* transgene in single-copy T2 *A. thaliana* plants (pooled non-vernalized whole plant samples from all lines per construct). The first nested PCR is shown. Spliceoforms were sequenced and categorised according to *A. thaliana* classes and Fig. 4.7. (B) Nucleotide sequences for the region pertaining to the *COOLAIR* distal splice site for each construct are given, and beneath that all detected transcripts. Potential splice sites are underlined, actual splice sites are in bold and mutations are highlighted in blue. (C) H4 to H5 and the first r-turn of the predicted *COOLAIR* structure for the wildtype *FLCA3a* (A3a.0) construct is given (left), and a possible alternative structure it could form (right). Outlined regions represent the segments that would be missing in the three alternate spliceoforms detected in the mutant (A10.1, A10.2, A10.3) constructs.

The *BrFLCA3a* mutant constructs did not express the desired II.iv (2) transcript, but the altered splice forms were all sufficient to significantly disrupt secondary structure (Fig. 5.11C). The shared A3a.3 and A3a.2 transcript is only 4 bp shorter than the canonical *BrFLCA3a* transcript and yet even this would shorten H4 to 3-6 bp, if not prevent its formation altogether. The A3a.1 and shared A3a.1 and A3a.2 transcripts are sufficient to remove H4 and consequently give rise to the predicted structure in Fig. 5.8 This is true for

both the predicted PTM *BrFLCA3a* secondary structure, with shorter H4 and altered first r-turn (left), and a possible alternative where sequence shifts from H5 to retain a H4 length and structure more like that from the R018 *BrFLCA3a* allele (right). Although we have created *BrFLCA3a* variants with altered distal transcription and secondary structure, there were no differences in flowering time at T2. It may be that the mix of hemi- and homozygous *FLC* alleles across the 10 representative plants for each line concealed phenotypic differences. This could be checked at T3 but seems unlikely as variation within lines was low. The most straight-forward explanation is that these changes at *COOLAIR* did not affect *FLC* sense expression or flowering time. In Chapter 4, we suggested from secondary structure analysis that the distal *COOLAIR* transcript from the PTM *FLCA3a* allele was a weak regulator of the sense; these experiments suggest it is non-functional. This would explain why there is no difference between wildtype and variant lines, despite changes in transcription and secondary structure. *FLCA3a* constructs were not analysed any further due to the lack of flowering time phenotype. It would, however, be interesting to test whether there are indeed functional differences between the distal *COOLAIR* transcripts expressed from the *FLCA3a* allele found in R018 versus that found in PTM.

5.2.4 A distal splice site mutation in FLCA10 *alters expression levels and flowering time*

Non-vernalized flowering times for *BrFLCA10* lines were recorded at T1 (Fig. 5.12A); non-vernalized and four weeks vernalized flowering times were recorded at T2 (Fig. 5.12B and Fig. S10). A10.1 and A10.3 lines are difficult to distinguish from A10.0 lines at T1, with a wide spread of flowering times (Fig. 5.12A). Differences in flowering may have been confounded by the presence of mixed copy numbers; however, there were no clear trends even when flowering time of single copy lines were analysed separately (not shown). In contrast, there is a marked difference in the flowering times of A10.2 versus wildtype A10.0 lines; the former had all flowered by 45 days, whereas many of the latter flowered after 45 days with some lines not flowering at all (in the 85 day trial). These trends continued at T2, with A10.2 lines flowering ~20 days earlier than A10.0, A10.1 or A10.3 lines (Fig. 5.12B and Fig. S10). All four constructs responded to vernalization. A10.0, A10.1 and A10.3 lines flowered 10-20 days earlier after four weeks vernalization. Some lines were not fully vernalized after four weeks and so exhibited a wider spread of flowering times, i.e. A10.0 line 58. A10.2 lines were early flowering under non-vernalized conditions but still exhibited a vernalization response, flowering ~10 days earlier after four weeks cold. A10.2 vernalized flowering times were uniform; this suggests vernalization requirement was saturated.

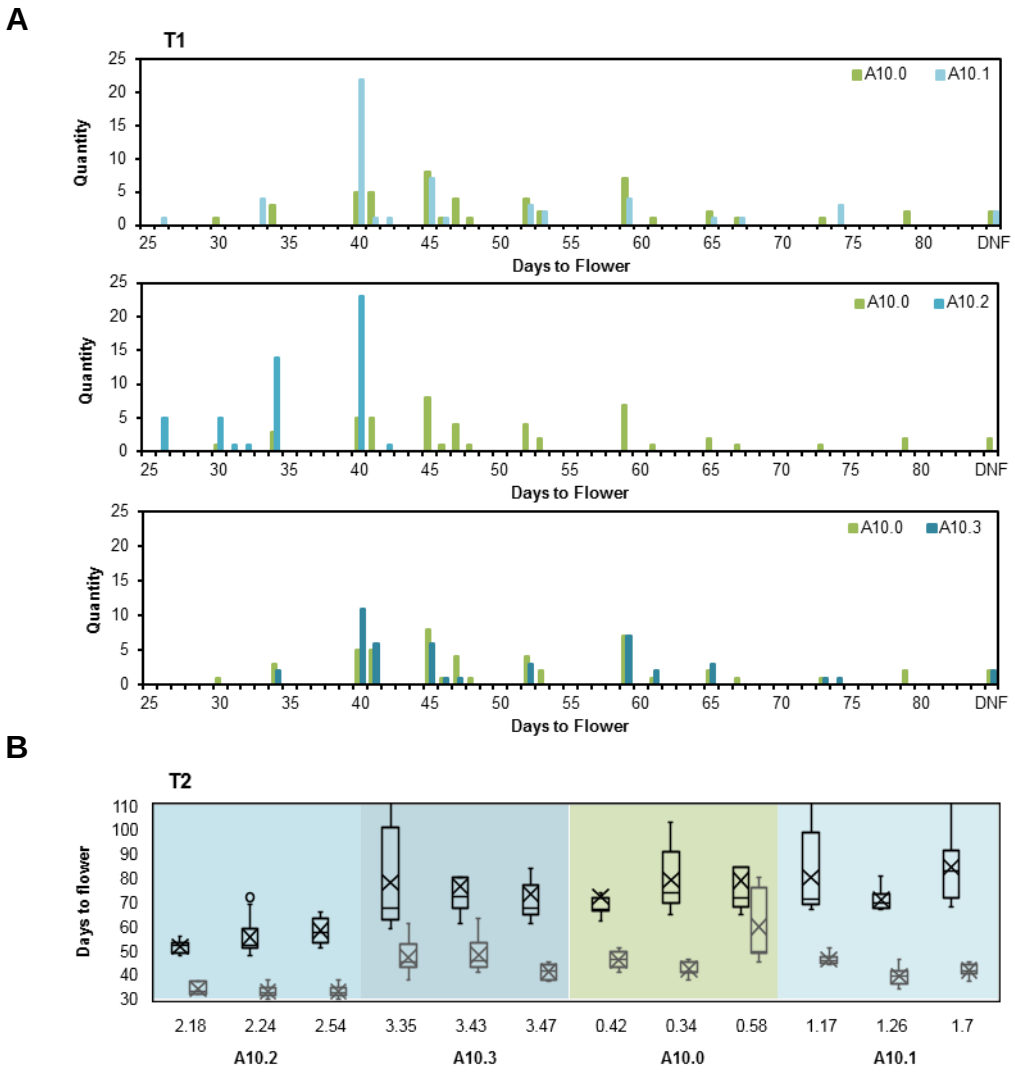


Figure 5.12: Flowering times of *A. thaliana* transgenic plants with *BrFLCA10* wildtype and mutant transgenes. (A) Flowering times for non-vernalized T1 multi-copy lines from each mutant construct directly compared with multi-copy lines from the wildtype construct: A10.1 versus A10.0 (top), A10.2 versus A10.0 (middle), A10.3 versus A10.0 (bottom). (B) Non-vernalized (black) and vernalized (grey) flowering times for T2 single-copy lines, represented by three median lines from each construct (10 plants/line). 2.18 refers to construct A10.2 line 18, 2.24 to construct A10.2 line 24 etc. For T2 flowering times of all lines see Fig. S10. Note that A10.0 is the wildtype *FLCA10* construct from *B. rapa* PTM; A10.1, A10.2 and A10.3 are mutant constructs.

T2 plants were vernalized for four weeks. Individual (for each line) and pooled (for each construct) samples were collected after 10 days in non-vernalized conditions, and at T0, T10 and T30 time-points after four weeks vernalization. *COOLAIR* distal splicing was analysed by the developed method (touchdown nested semi-quantitative RT-PCRs) for each of the constructs from pooled non-vernalized and V4WT0 samples. So far, only *FLCA10*.3 *COOLAIR* distal isoforms have been successfully detected (Fig. 5.13A). Expression of distal and proximal *COOLAIR* transcripts were confirmed in all four constructs through sequencing of short qPCR primer products. The developed *COOLAIR* splicing detection method may therefore not be working because of unforeseen changes in the *FLCA10* transcripts in *A. thaliana* or because of low expression levels. Alternatively,

the method developed for *COOLAIR* detection from *FLCA10* in *Brassica* may be unsuitable in *A. thaliana* and should be refined. A10.3 distal *COOLAIR* transcripts use a single splice site, which is the same as the wildtype *BrFLCA10* (and Var2-6) distal splice site. Despite mutations designed to reinstate the Col-0 splice site, the Var2-6 splice site continues to be used and the distal transcripts remain unaltered. Consequently, there was no change in flowering time. It is likely that A10.1, with similar flowering times to A10.0 and A10.3, also uses the canonical splice site. This suggests that reinstating the Col-0 site is not enough to switch usage from the Var2-6 site, despite previous evidence that the former is stronger. This could be because of the unique sequence context of *BrFLCA10*. The A10.2 construct was designed in case of this scenario: it has a mutated Var2-6, in addition to a reinstated Col-0, splice site. This prevents use of the Var2-6 site, although does not necessitate use of the Col-0 splice site – a different site altogether could be used. Interestingly, A10.2 lines flower significantly earlier than the other constructs. This suggests that these transcripts are indeed spliced differently. Splicing analysis of the A10.2 construct is required to confirm this.

Spliced *FLC* sense expression levels are ~2-fold lower in A10.2 pooled lines than in wildtype (A10.0) pooled lines, under non-vernalized conditions (Fig. 5.13B). Reduction in the number of *FLC* mRNA transcripts equates to reduction in FLC protein; this is the likely cause of the earlier flowering times described above. *FLC* unspliced expression is a useful indicator of changes in transcription. *FLC* unspliced expression mirrors that of spliced, with ~2-fold lower expression in A10.2 than wildtype (A10.0) lines. This suggests that the polymorphisms in A10.2 affect *FLC* sense transcription rather than RNA stability. *FLC* spliced and unspliced expression levels in A10.1 and A10.3 pooled lines are similar to the wildtype (A10.0) (Fig. 5.13B). *FLC* spliced and unspliced expression decreases in all four constructs after four weeks vernalization. A10.0, A10.1 and A10.3 spliced *FLC* expression decreases by about half in the cold, remains low 10 days after return to the warm, but then reactivates by 30 days. A10.0, A10.1 and A10.3 vernalized plants flowered earlier after four weeks cold than non-vernalized plants, but flowering time was still later than Col *FRI flc-2* plants. These data suggest that vernalization requirement was not saturated. A10.2 spliced *FLC* expression undergoes a small (~25%) decrease during cold, with levels remaining stable after return to the warm. This matches flowering time data, where we observe faster flowering times in plants with the A10.2 construct after cold treatment.

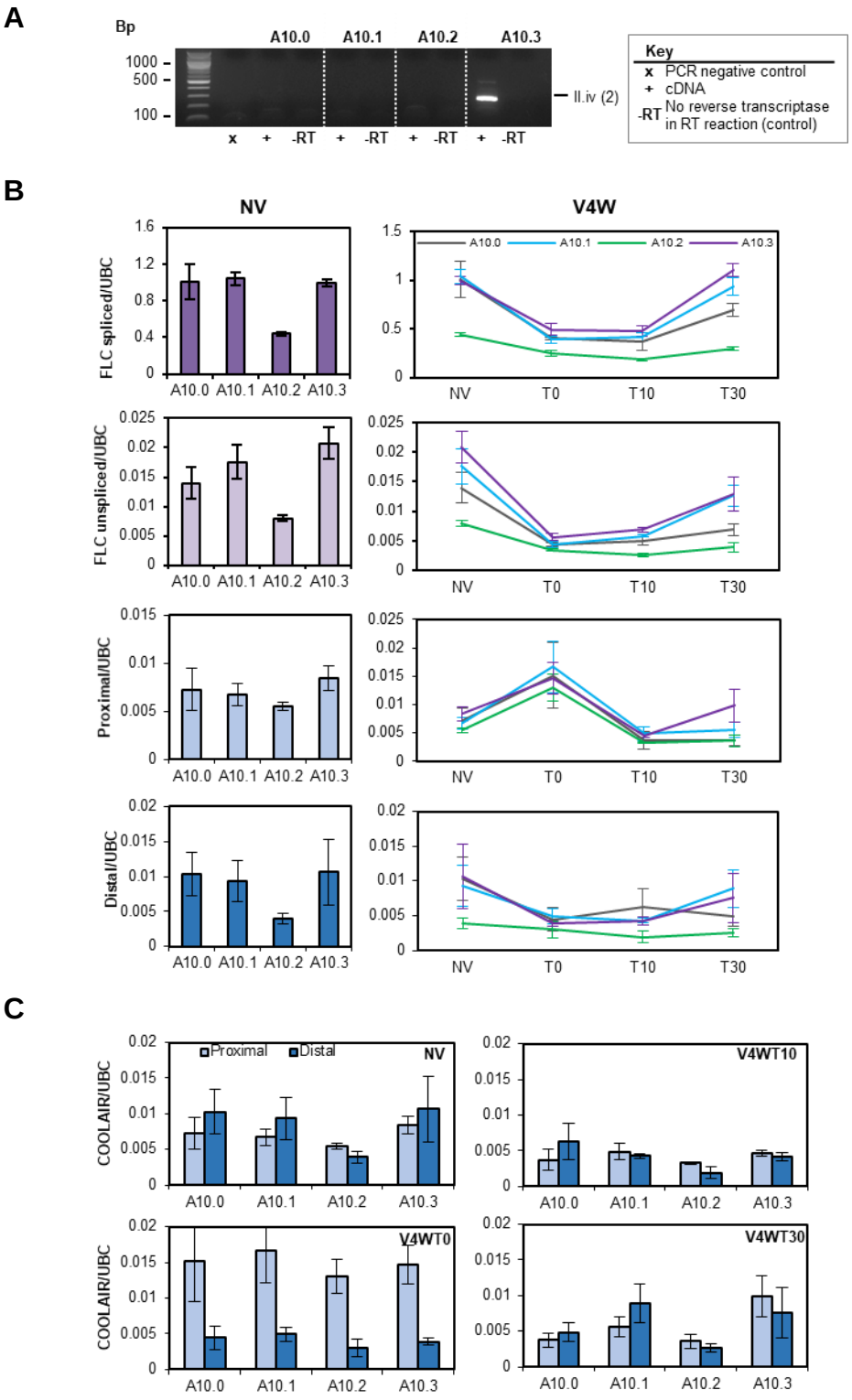


Figure 5.13: Sense and antisense expression patterns from the *BrFLCA10* wildtype and mutant transgenes in *A. thaliana*. (A) Touchdown nested RT-PCRs to detect distal *COOLAIR* transcripts from the *BrFLCA10* transgene in single-copy T2 *A. thaliana* plants (pooled non-vernalized whole plant samples from all lines per construct). The first nested PCR is shown. Spliceoforms were sequenced and categorised according to *A. thaliana* and *Brassica* classes (Fig. 3.1 and Fig. 4.7). *COOLAIR* was only successfully detected from A10.3 in this experiment. (B) Expression of *FLC* sense and antisense transcripts for each of the four constructs from non-vernalized (NV, left) whole plant and 4 weeks vernalized (V4W time-series, right) whole plant or leaf tissue (T2 *A. thaliana* transgenic plants). Whole plant samples for the time-series (right) were collected at 0 and 10 days after vernalization, and leaf samples at 30 days. Samples were normalized to the *UBC* reference gene; error bars are 1 St Dev. from mean for 2-3 biological replicates. (C) Results as above, but a direct comparison of antisense proximal and distal expression levels at the four time-points.

Unspliced *FLC* expression patterns during and after cold are similar to spliced. Unspliced *FLC* does not appear to reactivate by as much 30 days after cold; levels are lower rather than equal to non-vernalized levels. A10.1 and A10.3 may also reactivate more than A10.0 transcripts; this is most apparent from the *FLC* unspliced data. It does not translate into flowering time differences in the vernalized lines. Mutations in the A10.1 and A10.3 constructs are in the nucleation region and therefore may disrupt PHD-PRC2 accumulation in the cold. A10.2 contains the same mutation at the Col-0 site as A10.1; in this construct, any changes in nucleation are likely veiled by the decrease in non-vernalized *FLC* sense transcription.

Proximal *COOLAIR* expression patterns are similar across the four constructs under non-vernalized and vernalized conditions (Fig. 5.13B). Proximal *COOLAIR* increases up to 3-fold during cold, decreases to below original levels 10 days after cold and then remains low, mirroring patterns observed in *A. thaliana* (Swiezewski *et al.*, 2009). The decrease to below original levels likely reflects subsequent epigenetic shut-down of the *FLC* locus. Possible reactivation of A10.3 proximal *COOLAIR* transcripts (to non-vernalized levels) may reflect reactivation of this *FLC* construct 30 days after cold treatment. These data suggest that polymorphisms between the four constructs do not affect proximal transcription, as expected. Distal *COOLAIR* expression is lower in A10.2 than in the wildtype, A10.1 and A10.3 constructs and is unaffected by cold. Interestingly, during cold, distal *COOLAIR* expression decreases in the wildtype construct (and in A10.1 and A10.3), before stabilising (or even reactivating). In *Arabidopsis thaliana*, expression of both proximal and distal transcripts increase in the cold, although the proximal is induced to a greater extent (Csorba *et al.*, 2014). This phenomenon is perhaps unique to *BrFLCA10* or to the *Brassica FLC* homologues.

The relationship between proximal and distal *COOLAIR* appears to have been disrupted in the A10.2 construct (Fig. 5.13C). Bearing in mind large error bars (1 standard deviation from the mean) for distal expression, there is apparently a higher ratio of distal to proximal *COOLAIR* transcripts under non-vernalized conditions in the wildtype (and A10.1/A10.3) lines. Again, this is interesting because, to the best of our knowledge, it has not previously been observed in *A. thaliana*. This is reversed in the A10.2 early flowering lines; there is a higher ratio of proximal to distal transcripts, better representative of *A. thaliana* Col-0. A high proximal:distal *COOLAIR* ratio is associated with a less active *FLC* and earlier flowering in *Arabidopsis* before and after cold (Liu *et al.*, 2010; Csorba *et al.*, 2014). In all four *BrFLCA10* constructs, the proximal:distal ratio increases in the cold, suggesting

COOLAIR is involved in *BrFLC* transcriptional shut-down in the cold via a similar mechanism as for *AtFLC*. A return to a higher ratio of distal to proximal transcripts after cold (for all constructs except A10.2) correlates with reactivation of the sense. We suggest a link between the altered proximal:distal ratio in A10.2 and its early flowering phenotype. This experiment should be repeated before drawing any firm conclusions due to large error bars.

5.2.5 **COOLAIR** *splice site mutations at* **FLCC2** *do not affect flowering phenotype*

Different *COOLAIR* proximal and distal spliceoforms from *FLCC2* are expressed in winter versus summer lines, and in cold versus warm conditions. It was hypothesised that (1) different spliceoforms are stronger or weaker regulators of the sense, and (2) that the ratio of one spliceoform to another would therefore affect flowering time. To test this, we designed experiments to block expression of one (or more) spliceoforms and thereby promote expression of other forms. *BoFLCC2* transgenes with different splice site mutations were transformed into *A. thaliana* with a defective *FLC* gene (Col *FRI flc-2*) to negate compensation from other *FLC* homologues in *B. oleracea*. The late flowering *BoFLCC2* E9 allele from Purple Sprouting Broccoli (PSB) described by Irwin *et al.* (2016) was used for this purpose; this was the cultivar from which *BoFLCC2 COOLAIR* expression was initially detected in Chapter 4. A wildtype *BoFLCC2* construct from PSB had previously been designed and transformed into *A. thaliana* by Clare Lister and Judith Irwin (Irwin *et al*., 2016). Non-vernalized plant material from lines 5 and 266 from their study were used to confirm *COOLAIR* expression was the same in *A. thaliana* as in *B. oleracea* (Fig. 5.14A). The proximal Class I.ii form, but not Class I.i, was detected; similarly, Class I.i is not detected before cold treatment in *B. oleracea*. All three distal II.i constructs were detected, in addition to two Class II.iii transcripts. Class II.i (4) was only detected in *B. oleracea* after cold treatment, whereas it was detected in the *A. thaliana* transgenic lines in the warm. *FLCC2* constructs from *B. oleracea* PSB were designed with mutations to block splice sites at P1, P2, D1, D2 or D3 and so promote the production of alternate isoforms (Fig 5.14B). The five variant constructs contain one SNP, each at a different splice site. It is predicted that this will promote transcription of the other spliceoforms; however, it is possible that an alternate up- or down-stream splice site will be used (underlined in Fig. 5.14C). For example, the C2. D1 construct is designed to block use of the ‘D1’ 3’ splice acceptor site, preventing canonical splicing of II.i (2) and II.i (4) transcripts and promoting transcription of the II.i (3) transcript. The C2.D3 construct blocks use of the D3 5’splice acceptor site, preventing canonical splicing of II.i (4), and promoting II.i (2) and (3) transcripts.

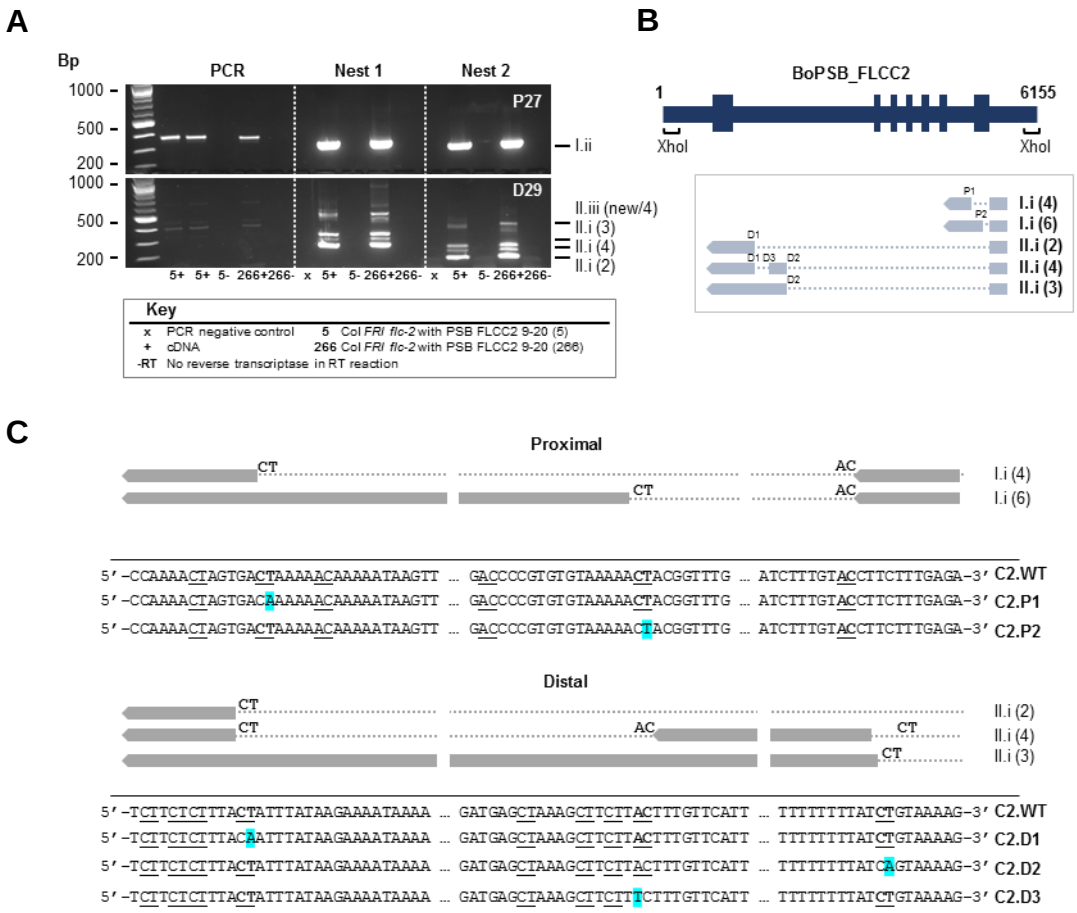


Figure 5.14: *Brassica FLCC2* construct design (A) Touchdown nested RT-PCRs to detect distal *COOLAIR* transcripts from *BoFLCC2* E9 constructs in *A. thaliana* transgenic lines 5 and 266. These constructs were designed, synthesised and transformed into *A. thaliana* by Clare Lister and Judith Irwin and used with permission (Irwin *et al*., 2016). Spliceoforms were sequenced and categorised according to *A. thaliana* and *Brassica* classes (Fig. 3.1 and Fig. 4.7). (B) Schematic of the *B. oleracea* PSB *FLCC2* construct and the spliceoforms that are expressed in the wildtype construct (also see Fig. 4.7). Splice sites targeted for mutation are labelled P1, P2, D1, D2 and D3 (which also represent their mutant construct names). (C) Nucleotide sequences of the proximal (top) and distal (bottom) targeted *COOLAIR* splicing regions in the four constructs. Potential antisense splice sites are underlined, canonical *A. thaliana* sites are in bold, and mutations highlighted in blue. '…' represents additional sequence that is not shown.

The five *BoFLCC2* variant constructs were synthesised via site-directed mutagenesis of a pBlueScript plasmid containing the wild-type *Brassica FLCC2* E9 allele (Irwin *et al*., 2016; Fig. 2.3). The six constructs (wildtype plus five variants) were cloned into pSLJ1714 for transformation into Col *FRI flc-2* plants via *Agrobacterium*-mediated floral dip (Table 5.2). Positive T1 transformants were identified by Kanamycin selection. There were only 10 positive wildtype (C2.WT) transformants, of which one died before seed set and only two contained a single copy of the construct. C2.9-20 T2 plants (Irwin *et al*., 2016) containing the same *FLCC2* wildtype construct were therefore used to supplement the two single-copy C2.WT plants at T2. Copy number analysis was performed from leaf tissue DNA at T1 stage for the six original constructs, and at T2 for the 9-20 constructs (as copy

numbers had previously not been checked) (Fig. 5.15A). Only single copy T2 9-20 plants were included in the expression analysis and flowering time data.

ID	Construct	Vector	Background	T1 Lines	T2 Lines	T2 Plants
C2.WT	*B. oleracea* PSB E9 FLCC2 WT	pSLJ1714	*A. thaliana* Col *FRI flc-2*	10	2	10
C2.P1	*B. oleracea* PSB E9 FLCC2 v1	pSLJ1714	*A. thaliana* Col *FRI flc-2*	117	15	10
C2.P2	*B. oleracea* PSB E9 FLCC2 v2	pSLJ1714	*A. thaliana* Col *FRI flc-2*	64	15	10
C2.D1	*B. oleracea* PSB E9 FLCC2 v3	pSLJ1714	*A. thaliana* Col *FRI flc-2*	68	16	10
C2.D2	*B. oleracea* PSB E9 FLCC2 v4	pSLJ1714	*A. thaliana* Col *FRI flc-2*	75	15	10
C2.D3	*B. oleracea* PSB E9 FLCC2 v5	pSLJ1714	*A. thaliana* Col *FRI flc-2*	60	15	10
C2.9-20	*B. oleracea* PSB E9 FLCC2 WT	pSLJ1714	*A. thaliana* Col *FRI flc-2*	N/A	12	10

Table 5.2: *BoFLCC2* construct details. Number of lines analysed per construct at T1 and T2 stage and the number of plants analysed per line at T2.

Flowering times were analysed at T1 after 4-12 weeks vernalization. Selection of constructs with low numbers of positive transformants had to be repeated several times, hence the range of vernalization times (Fig. 5.15B). 7-68 lines from each construct experienced 12 weeks vernalization and therefore flowering times from these are compared at T1 (Fig. 5.15C left). Allowing for disparate representation of constructs and a wide spread of flowering times, there were small differences in mean flowering time of each construct. The five variant constructs appeared to flower earlier than C2.WT by 5-10 days. C2.P1 constructs also appeared to flower slightly later (by ~5 days) than C2.P2 constructs. These differences might be underpinned by non-optimal *COOLAIR* ratios. C2.WT is, however, severely underrepresented in this set compared with the other constructs, and we therefore may just be observing genome position effects. When flowering times from the other vernalization sets are added, we note that there is still a wide range of flowering times with flowering not obviously altered between four and 12 weeks vernalization (Fig. 5.15C right). This suggests vernalization requirement is not saturated in many of these lines even after 12 weeks cold. Flowering times were analysed at T2 from non-vernalized and 10 weeks vernalized single copy lines (Fig. 5.15D). At T2 there were no obvious differences in flowering time between constructs under non-vernalized conditions or after 10 weeks cold, which appeared to saturate vernalization requirement. Further analysis at T3 would help to determine whether small differences have been disguised by the mix of homo- and hemizygous alleles for each construct at T2. Some *BoFLCC2 COOLAIR* transcripts were cold-induced; it would therefore also be useful to check T3 flowering time differences before cold saturation to analyse vernalization response. Otherwise, these data suggest that the mutations have either not sufficiently altered *COOLAIR* ratios, or that it is the presence or absence of *COOLAIR*, rather than the presence/absence of specific transcripts, that is primarily important for flowering time regulation.

A

Copy number	C2.WT	C2.P1	C2.P2	C2.D1	C2.D2	C2.D3	C2.9-20
1	2	42	24	23	39-40	26	12
2	4	33	12	15	12-13	16	4
3	1	11	9	10	6	6	4
4	0	7	6	6	2	2	1
5	0	5-7	2	5	2	1	0
6	2	3-5	2	1	2	1	1
7	0	1	0	1	0	0	0
8	0	1	0	0	2	0	0
9	0	0	0	1	0	0	0
10	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0
12	0	1	0	0	0	1	0

B

T1 plants (no)	Vern. (weeks) 12	8	6	4
C2.WT	7		2	1
C2.P1	22		35	60
C2.P2	64			
C2.D1	68			
C2.D2	15	60		
C2.D3	60			

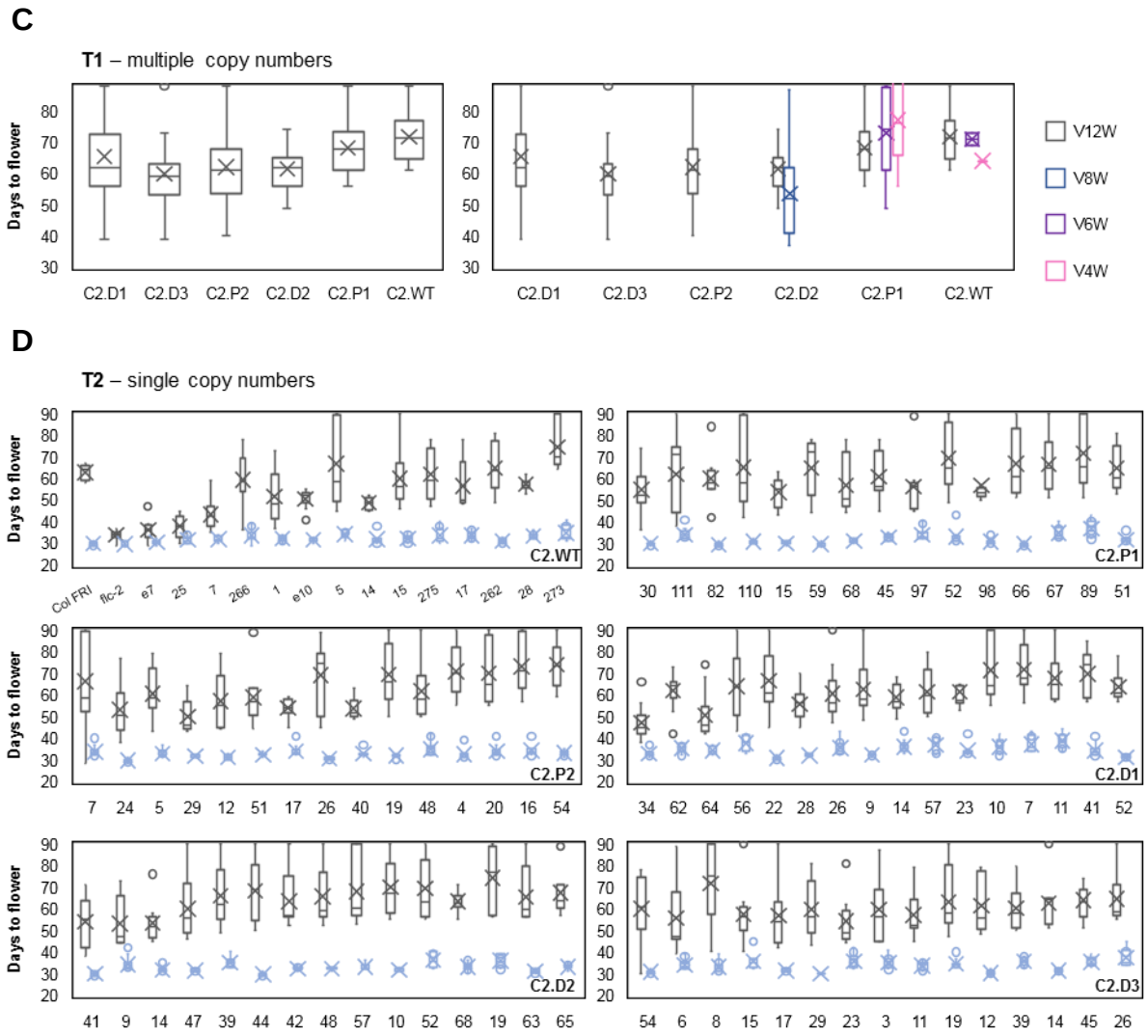


Figure 5.15: Flowering times of *A. thaliana* transgenic plants with *BoFLCC2* wildtype and mutant transgenes. (A) Copy numbers of *BoFLCC2* transgenes after *Agrobacterium*-mediated floral dip at T1, single copy lines were carried forwards to T2. (B) Table with vernalization times for each construct at T1. (C) Average flowering times for T1 multi-copy lines from each construct after 12 weeks vernalization (left) and for all vernalization time periods (right). (D) Non-vernalized (grey) and vernalized (blue) flowering times for T2 single-copy lines for each construct, and for the *A. thaliana* Col *FRI* and *A. thaliana* Col *FRI flc-2* controls (see C2.WT); 10 plants were analysed per line. Note that C2.WT is the wildtype *BoFLCC2* construct from *B. oleracea* PSB; C2.P1, C2.P2, C2.D1, C2.D2 and C2.D3 are mutant constructs.

5.3 Discussion

5.3.1 Natural variation can alter COOLAIR *structure and processing in* Brassica

Natural variation in flowering time has emerged primarily through polymorphisms in non-protein-coding regions of *FLC* (Li *et al.*, 2014). We are interested in those that disrupt normal processing or expression of *COOLAIR*, such as the splice site mutation in the Var2-6 accession which alters distal splicing (Li *et al.*, 2015). Rapid accumulation of polymorphisms in long non-coding (lnc)RNAs such as *COOLAIR* could provide a mechanism for positive selection (Pang *et al.*, 2006; Kutter *et al.*, 2012; Nitsche *et al.*, 2015). *COOLAIR* has been conserved in other members of the Brassicaceae family, including *A. lyrata*, *A. alpina*, *C. rubella*, *B. rapa*, *B. oleracea* and *B. napus* (Chapter 4; Castaings *et al.*, 2014; Hawkes *et al.*, 2016; Li *et al.*, 2016). We speculate that *COOLAIR* has been conserved as a natural tool to fine-tune flowering. Whereas multiple loss-of-function mutations at *FRIGIDA* have arisen independently to promote early flowering (Shindo *et al.*, 2005), we are uncertain how many phenotypically relevant mutations work through *COOLAIR*. So far, Var2-6 is the only experimentally-validated example, with the contribution of altered *COOLAIR* polyadenylation to the Lov-1 phenotype requiring further investigation (Coustham *et al.*, 2012). Other studies which link *COOLAIR* processing with flowering time are based on mutagenesis screens rather than natural populations (Liu *et al.*, 2010; Sun *et al.*, 2013; Marquardt *et al.*, 2014; Wang *et al.*, 2014A). Within the polyploid and paleopolyploid *Brassica* species, the presence of multiple *FLC* copies arguably provides greater opportunity for the accumulation of *cis* mutations due to redundancy between copies (Song *et al.*, 1995; Walsh, 1995; Osborn *et al.*, 2003; Schiessl *et al*., 2014). *COOLAIR* is transcribed from at least three of four *FLC* clades in *B. rapa* and *B. oleracea* (Chapter 4), although it is not known whether they regulate sense transcripts in the same way. Here, we tested how distinct *Brassica COOLAIR* isoforms regulate expression of their corresponding *FLC* homologue.

Genomic plasticity of the polyploid *B. napus* is thought to have the facilitated both large and small modifications in flowering time between cultivars (Schiessl *et al*., 2014). We identified natural polymorphisms at *FLC* between winter and spring *B. napus* cultivars, including non-synonymous SNPs, non-coding SNPs, insertion-deletions, and, for *BnFLCC2*, presence-absence variation. Similarly, considerable sequence and copy number variation was identified between *FLC* alleles in four diverse *B. napus* morphotypes: winter oilseed, spring oilseed, fodder rape and swede (Schiessl *et al*., 2014). We divided the *BnFLC* sequences into nominal alleles, akin to the division of *Arabidopsis FLC* alleles into

haplotype groups (Li *et al.*, 2014). This is less straightforward in *Brassica*, given that intra-plant *FLC* variation (between homologues) is far greater than intra-species variation. *FLC* haplotype groups in *Arabidopsis* have unique patterns of geographical distribution, representative of the diverse environmental conditions they have adapted to (Li *et al.*, 2014). Natural variation is complicated in a polyploid, where a fast vernalizing *FLC* allele at one locus may be counteracted by a slow vernalizing *FLC* allele at another, and has further been exaggerated in *B. napus* by marker-assisted selective breeding and TILLING to meet agricultural demand (Jung and Müller, 2009). We identified *FLC* alleles which occur more frequently in winter types and alleles which occur more frequently in spring types; these may represent strong and weak alleles, respectively. Actual flowering times for winter and spring lines that contain 'strong' or 'weak' alleles are likely to vary because of differences at other *FLC* copies (or at other flowering time regulators). S31, for example, contains a 'weak' *BnFLCA3a* allele and a 'strong' *BnFLCA10* allele and yet flowers early. Previously reported protein-coding polymorphisms at *FLCA10* were detected in this set; for example, the G to A SNP in *FLCA10* intron 6 that disrupts sense splicing (Yuan *et al.*, 2009). We further observed a non-coding C to T SNP in the second RY motif of the 'strong' *FLCA10* allele; this may inhibit VAL1-binding and delay epigenetic silencing via a mechanism akin to the *vrn8* mutation in *Arabidopsis* (Qüesta *et al.*, 2016). This demonstrates how our extensive knowledge of flowering time processes in *Arabidopsis* can be used to improve characterisation of flowering time in *Brassica*.

If antisense transcription at *FLC* finetunes flowering time, we might expect natural variation to have arisen through *COOLAIR* multiple times in evolutionary history, as observed for *FLC* sense transcripts (Michaels *et al.*, 2003; Yuan *et al.*, 2009; Wu *et al.*, 2012; Zou *et al.*, 2012). We identified polymorphisms which affect *COOLAIR* processing and/or expression in *BnFLCA3a* and *BnFLCC2*. Their occurrence in two out of the three tested *FLC* homologues supports this as a possible mechanism for natural variation. Most interesting are the SNPs that instate or block a *BnFLCC2* proximal small intron splice site. The major form of the proximal that is transcribed in the early flowering spring cultivars is the Class I.ii unspliced variant. *COOLAIR* proximal processing is cotranscriptionally coupled with *AtFLC* gene body histone methylation and sense transcription (Marquardt *et al.*, 2014). Paradoxically, the *AtFLC* unspliced proximal transcript correlates with high sense expression and delayed flowering. *Cis*-polymorphisms at the 3' end of the *BnFLCA3a* proximal transcript could also prove interesting, given that the SNP at the 3' end of the distal transcript in Lov-1 has been associated with increased distal polyadenylation (Coustham *et al.*, 2012). We do not yet know if these polymorphisms are

responsible for phenotypic differences; association studies would be a useful way to test this (Jung and Müller, 2009).

Polymorphisms which alter splicing will have associated consequences for secondary structure; other polymorphisms can impact secondary structure directly. A key example, is the *COOLAIR* intra-exon SNP between PTM and R018 *BrFLCA3a* alleles, predicted to shorten the distal helix H4 in the former. It is difficult to predict RiboSNitches – SNPs that significantly disrupt structure – manually or by computer modelling (Ritz *et al.*, 2012; Corley *et al.*, 2015). Structural predictions in Chapter 4 suggest that *COOLAIR* helices are more often maintained than not, despite multiple mutations. A mutation must not only disrupt a base pair, but also favour the switch from one network of base pairing and stacking interactions to another (Ritz *et al.*, 2012). A SNP that reduces a double-stranded segment of a helix to two or less base pairs is more likely to disrupt that helix by expanding internal loops. Even then, functional consequences depend on whether the disrupted region is involved in regulatory interactions. The predicted PTM RiboSNitch should therefore be validated by chemical-probing.

These experiments establish (1) the usefulness of sequencing short regions of *FLC* pertaining to *COOLAIR* to identify potentially significant polymorphisms and (2) that non-coding *cis* polymorphisms in *Brassica* can alter *COOLAIR* splicing and structure. This confirms that mutations that disrupt canonical *COOLAIR* processing have arisen more than once over evolutionary time, but we do not yet know whether they contribute to positively selected for phenotypes.

5.3.2 Induced variation can alter COOLAIR *structure and processing in* Brassica

In Chapter 4, differential splicing and expression patterns of *COOLAIR* were linked to natural variation between *FLC* homologues. We further observed differential expression of *COOLAIR* spliceoforms under different environmental conditions and between winter versus spring cultivars. In this chapter, transgenic experiments were set up to test whether natural variation in *COOLAIR* structure and processing could alter *cis FLC* expression and flowering time phenotype.

Experiments were designed to block or promote specific proximal and distal isoforms through single mutations at *BoFLCC2 COOLAIR* splice sites, akin to SNPs that could have arisen naturally over time. These mutations had no obvious effect on flowering time (*FLC* sense expression levels could be checked at T3 to confirm this), signifying that they were

not sufficient to alter *COOLAIR* regulation of the sense. This might be because the different splice forms act redundantly, or because the changes were not sufficient to significantly disrupt *COOLAIR* splicing or regulation; splicing in the transgenic lines should therefore be checked. These data suggest natural selection through *COOLAIR* could be rarer than hypothesised. Parallel Analysis of RNA Structure has revealed a multitude of riboSNitches in humans, many of which are associated with human disease phenotypes (Martin *et al.*, 2012; Wan *et al.*, 2014). Certain human haplotypes are thought to be conserved to reduce the probability of riboSNitches (Martin *et al.*, 2012). To the best of our knowledge, it is not known whether (or how many times) different riboSNitches have evolved at the same RNA to alter phenotype over evolutionary time. We suggest that (1) a threshold number of mutations, and/or (2) mutations localised to specific regions (of functional significance) are required to disrupt *COOLAIR*.

Greater variation exists between *Brassica FLC* homologues within a plant than between alleles across a species. This has been driven by chromosome reshuffling, divergence over evolutionary time, and lower sequence constraint in the *Brassica* paleopolyploids (Walsh, 1995; Lynch and Conery, 2000; Osbourn *et al.*, 2003; Cheng *et al*., 2013). We therefore have biological examples of *COOLAIR* species with a large number of polymorphisms, indels and/or sequence shifts. These are distinct in underlying sequence, splicing and structure, and perhaps also in their functional behaviour (Chapter 4). From previous work, we suggest that the structural helix H4 and first r-turn domains are key to *COOLAIR* function (Li *et al.*, 2015; Hawkes *et al.*, 2016; Chapter 3 and 4). A low number of mutations localised to this region may therefore be sufficient to alter function. We tested this through transgenic experiments to switch *Brassica rapa FLCA10* and *FLCA3a* canonical splice sites and, consequently, length and/or presence of H4 and the first r-turn.

The alternate distal transcripts transcribed from the three *BrFLCA3a* variant constructs were predicted to significantly disrupt H4 and the first r-turn. Despite this, we observed no differences in flowering time in T2 populations under non-vernalized conditions or after four weeks cold. We tentatively propose that the distal transcript from this *BrFLCA3a* allele (from the PTM winter cultivar) was already significantly disrupted by the previously identified SNP in helix 4. This would explain why subsequent splicing mutations had no effect. A comparative functional analysis of *COOLAIR* transcripts from the R018 *BrFLCA3a* allele should be performed to validate this. Earlier flowering times were observed with the A10.2 variant of the wildtype *BrFLCA10* construct, where the normal splice site was mutated and a longer helix 4 predicted. This correlated with reduced

BrFLCA10 sense and distal *COOLAIR* expression, and comparatively higher levels of proximal expression. This suggests that the altered distal *COOLAIR* transcript was a stronger negative [or weaker positive] regulator of the sense than the canonical. Experiments to analyse splicing were unsuccessful and therefore we are currently uncertain whether the reinstated Col-0 site is used. Any alternate splice site 5' of the original site, however, is likely to increase helix 4 length. Perhaps somewhat paradoxically, we observe lower levels of the new 'strong' distal regulator in A10.2. One possibility is that the total number of spliced distal products are reduced because of low affinity between spliceosome components and the new splice site (Reed and Maniatis, 1986). Reduced amounts of a negative regulator, however, would likely cause an increase in *FLC* sense transcription. Instead, we propose that the strong distal transcript pushes *FLC* into a low activity state, leading to a build-up of repressive chromatin marks, slow transcription and, in turn, a higher proximal:distal ratio (Liu *et al.*, 2010; Csorba *et al.*, 2014). This is similar to the sense-antisense feedback mechanisms observed by Wang *et al*. (2014). These experiments support the hypothesis that differences in splicing and structure at *BrFLCA10* can affect *COOLAIR* regulation of the sense transcript in *Brassica*, as it does in *Arabidopsis*.

Further experiments are needed to prove that the two single nucleotide mutations in A10.2 work through *COOLAIR*. The A10.1 transcript acts as a control for one of these: it also contains the SNP that reinstates the Col-0 site but flowering time is unchanged. There is no control for the second SNP that mutates the Var2-6 site. This is because it was incorrectly assumed that the reinstated 'strong' Col-0 site would be reverted to in A10.1 or A10.3, without mutation of the Var2-6 splice site. Although the position of the second SNP (at the Var2-6 site) has not so far been associated with any regulatory mechanisms in *Arabidopsis*, it is highly conserved throughout species (Li *et al.*, 2015), is in close proximity to sense exon 1, and is within the nucleation region of *FLC*. Changes in the cold response mechanism are unlikely because sense transcript expression is reduced before cold; however, sense splicing should be checked. We could confirm whether this SNP works through *COOLAIR* by analysing the same mutation in a *BrFLCA10* construct that lacks *COOLAIR* transcription. Such a construct does not yet exist but this could, in the meantime, be tested in the *A. thaliana TEX* (Terminator EXchange) lines (Csorba *et al.*, 2014).

These experiments lend support to the idea that H4 and the first r-turn are functionally important structural domains. RNA structural domains can interact with protein and other nucleic acids to facilitate function (Brodersen *et al.*, 2002; Liu *et al.*, 2015; Somarowthu *et*

al., 2015; Xue *et al.*, 2016, and Lewis *et al.*, 2017 are just a few examples). One possibility is that the distal *COOLAIR* interacts with a H3K36 demethylase through its conserved r-turn, promoting the removal of active H3K36me3 chromatin marks (Csorba *et al.*, 2014). A similar 5' asymmetric internal loop in the *Braveheart* lncRNA was found to interact with a zinc-finger protein, facilitating regulation of cardiovascular lineage commitment (Xue *et al.*, 2016). The PTM *BrFLCA3a* allele is predicted to alter the angle of the r-turn and reduce H4 height in the distal transcript, making it non-functional. The PTM *BrFLCA10* allele is predicted to have a short or absent H4, also making it non-functional. Comparison of H4 length in Col-0 versus Var2-6 suggested that shortening of the helix quantitatively reduces functional ability (Hawkes *et al.*, 2016). The proposed effect of shortening distal H4 length on *FLC* expression and consequently flowering time is presented schematically in Fig. 5.16.

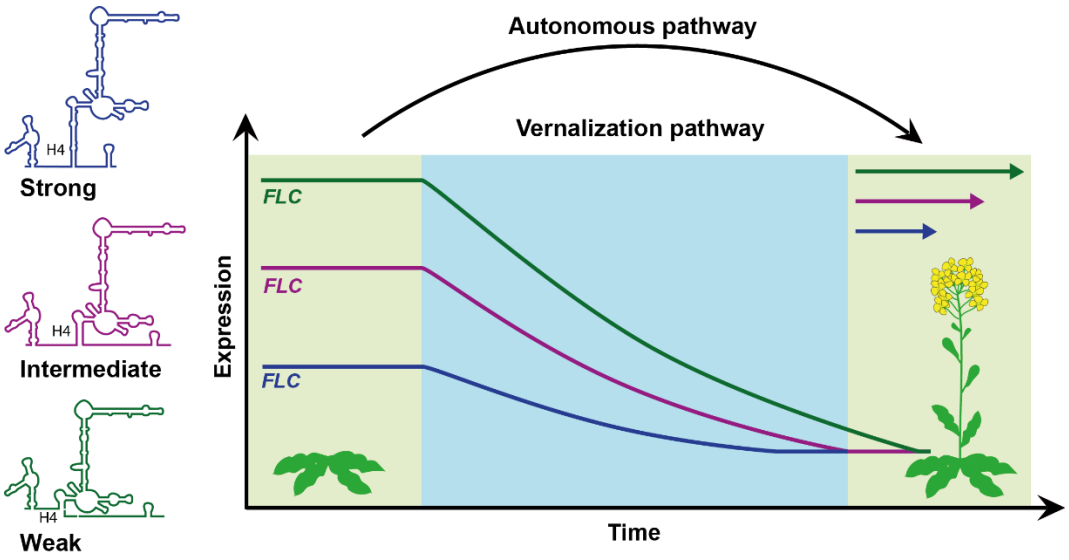


Figure 5.16: Predicted effect of shortening the distal *COOLAIR* helix H4 on *FLC* expression and flowering time. The green line represents *FLC* expression levels over time with a short H4; the pink line, *FLC* expression levels with an intermediate length H4; the blue line, *FLC* expression levels with a long H4. The green arrow represents longer flowering time with a short H4; the pink arrow, intermediate flowering time with an intermediate length H4; the blue arrow, shorter flowering time with a long H4.

Alternatively, there may be a cut-off point when, for example, the r-turn becomes inaccessible. Mutating the Var2-6 splice site to encourage use of an upstream site in the *BrFLCA10* A10.2 construct was predicted to lengthen H4, theoretically producing a stronger distal regulator. Accordingly, sense expression decreased and flowering time was earlier in A10.2 lines. It is puzzling that the 22 bp H4 of the Var2-6 distal Class II.iv (2) isoform is thought to have reduced function but is still longer or equal in length to those transcribed from the *B. rapa FLCA10*, *FLCA3a* and *FLCA2* homologues. This implies a

Brassica-specific reduction in *COOLAIR* distal regulation. Var2-6 also produces a Class II.iv (1) transcript, however, where H4 and the first r-turn are predicted to be completely absent (as there is no compensatory sequence from the extra exon) (Peijin Li, Dean lab, unpublished). It is therefore likely that both transcripts contribute to increased sense expression in Var2-6, not solely the Class II.iv (2) isoform analysed by Li *et al.* (2015) and Hawkes *et al.* (2016).

This work suggests that natural variation in *Brassica* flowering time can arise through mutations at *COOLAIR*, but that this is rare. Although mutations that alter *COOLAIR* splicing (and consequently structure) have arisen independently between different *FLC* homologues or *Brassica* cultivars, these do not necessarily impact its function. For example, different SNPs at *COOLAIR* in *BoFLCC2* did not appear to alter flowering time. Instead, we suggest that either a large number of mutations, such as those found between the *FLC* clades, and/or mutations in a specific structural domain, such as those synthesised in the A10.2 transgenic line, are necessary to alter *COOLAIR* behaviour. Such mutations give rise to a flowering time phenotype which can be selected for or against. These experiments support the idea that *COOLAIR* provides a tool to finetune flowering time, and that there is rich variety in that tool in *Brassica*. Its subtlety, however, is yet to be discerned. It may be that variation at *COOLAIR* has only occasionally led to flowering time differences over evolutionary time. Distinct *COOLAIR* variants are found at the different *FLC* homologues in *Brassica* and therefore likely play a role in expression differences between the homologues in *cis*.

5.3.3 Insights into splice site selection at COOLAIR

Splicing proceeds on nascent transcripts in a 5’ to 3’ direction, often within minutes of transcription (Beyer and Osheim, 1988). 5’-AG-3’ is a typical splice acceptor site in dicotyledonous plants; this reads 3’-CT-5’ for antisense transcripts (Hebsgaard *et al.*, 1996). A region 18-37 nt (or 20-30 nt in dicot plants) downstream of the branch site is typically scanned for suitable AG acceptor sites (Beyer and Osheim, 1988; Hebsgaard *et al.*, 1996). Mutations at *BrFLCA3a* successfully induced distal splice variants with downstream 3’ splice sites. The target Var2-6 site was not used, despite being conserved. In contrast, it is thought that splicing was only altered in one of three *BrFLCA10* constructs (A10.2); the primary Var2-6 site continued to be used in the two constructs where it was not mutated (A10.1 and A10.3). This is interesting in terms of what we learn about splice site dynamics.

Deletions, substitutions or even subtle changes in exon sequence can alter splice site selection (Reed and Maniatis, 1986). From the antisense 3'-CTG-5' to 3'-CTT-5' mutation in Var2-6, we suspect that the former is the stronger splice site. This is supported by distal splice site selection in the *BrFLCA3a* constructs; although the Var2-6 site is not used, other nearby 3'-CTG-5' sites are preferentially selected over 3'-CTH-5' sites. The nucleotide preceding the 5'-AG-3' splice site has been reported as critical in determining the competitiveness of closely spaced AG dinucleotides in mammals, with 5'-CAG-3' and 5'-UAG-3' (or antisense 3'-CTG-5' and 3'-CTA-5') sequences the strongest (Smith *et al.*, 1993). All selected *BrFLCA3a* splice sites in the mutated constructs were 5' of the Var2-6 site. When the strong downstream 3'-CTG-5' site was mutated (in addition to the Col-0 site) in A3a.3, splicing reverted to a weaker upstream 3'-CTT-5' site. Interestingly, it is only when the Col-0 site is left intact, but made weaker via the 3'-CTT-5' mutation, in A3a.1, that the site furthest downstream is used. This suggests that when there are fewer potential splice sites within the scanning region, the first available AG is more likely to be used. This may also explain why reinstating the Col-0 site in A10.1 and A10.3 did not necessarily promote its selection. This is supported by the scanning hypothesis or the first-come-first-served rule – that the first AG dinucleotide within 30 nucleotides of the branch site is used, assuming it is not very close to the latter or obscured by secondary structure (Reed and Maniatis, 1986; Beyer and Osheim, 1988; Smith *et al.*, 1993). In dicot plants, there is consequently strong selective pressure against multiple AG dinucleotides within the scanning region (Hebsgaard *et al.*, 1996); for the non-coding *COOLAIR* transcripts selection appears weaker.

A single distal isoform was transcribed from the wildtype A3a.0 construct, whereas two were transcribed from the mutated A3a.1 and A3a.2 constructs. Discrimination between cryptic and normal splice sites may result from the affinity of splicing components for these two types of sites (Reed and Maniatis, 1986). Less effective complex formation with cryptic sites may explain why multiple spliceoforms were transcribed when the canonical site was mutated in *BrFLCA3a*. Low spliceosome affinity may also contribute to reduced levels of a spliced transcript. Blocking of the normal 5' splice site of *IVS2* results in activation of a weak cryptic site and higher frequency exon skipping (Treisman *et al.*, 1982). This could contribute to lower levels of the distal *COOLAIR* transcript in the A10.2 *FLCA10* construct, in which the canonical splice site was mutated. Expression levels of the unspliced distal transcript should be checked in A10.2 to see if these are higher. We observed no variation in *BrFLCA3a* or *BrFLCA10* 5' splice sites, which was expected given that their immediate sequence context had not been changed.

6 Discussion

6.1 Decoding the non-coding genome

The Ensembl project began sequencing and annotating the human genome in 1999; it has since revealed that protein-coding genes comprise only ~1-2% of the genome, far lower than expected (Chi, 2016). In an affront to anthropocentrics, the number of protein-coding genes in humans has been predicted at around half that of a banana: ~19,000 versus ~36,500 (D'Hont *et al.*, 2012; Ezkurdia *et al.*, 2014). What then is the other 99% of the human genome doing? Humans (and all other eukaryotes) reportedly contain vast tracts of 'junk DNA' without biological significance or function, the remnants of evolutionary experiments built up over time. Mounting evidence, however, suggests that although not protein-coding, much of this 'junk DNA' is still transcribed (Clark *et al.*, 2011). Unravelling the 'dark matter' of the genome has gradually revealed the complex network of non-coding elements and RNAs that appear to regulate protein-coding genes (Chi, 2016). These non-coding elements perhaps provide the key to the evolutionary complexity of higher organisms (Mattick, 2011).

Complex functional mechanisms of many eukaryotic infrastructural and small RNAs have been characterised (Kim and Sung, 2012). In contrast, long non-coding (lnc)RNAs have proved far more enigmatic, not least because their categorisation as such depends only on their being over 200 nucleotides in length; they vary widely in genome positioning, processing, function and mechanism of regulation (Ponting *et al.*, 2009). There are multiple examples with experimentally-validated functions, including *Braveheart* (Xu *et al.*, 2016), *Chaer* (Wang *et al.*, 2016), *HOTAIR* (Tsai *et al.*, 2011), *MALAT1* and *TUG1* (Yang *et al.*, 2011), and *Xist* (Galupa and Heard, 2015); these have been discussed in greater detail in Chapters 1 and 3. Such studies, however, barely scratch the surface of the vast numbers of identified but uncharacterised long non-coding transcripts, leading to further debate over their functional significance. Estimates that 80% of the human genome is functional were ridiculed for not taking into account biological conservation, or the lack thereof (ENCODE, 2012 versus Graur, 2013). Indeed, many lncRNAs are less well-conserved than their protein-coding counterparts, indicating low conservation across the pool or some functionally significant RNAs amongst a background of transcriptional noise (Ponjavic *et al.*, 2007).

This thesis, in combination with a decade of research from the Caroline Dean lab, has helped to investigate the question mark hanging over lncRNA significance through analysis of the long non-coding antisense transcripts found at *FLC* in the reference plant, *Arabidopsis thaliana*. Chapter 3 and 4 demonstrate that *COOLAIR* has been transcriptionally and structurally conserved in five other members of the Brassicacea family, despite sometimes low primary nucleotide sequence identity. Evidence for structural conservation in two additional species is also presented; the evolutionary conservation of *COOLAIR* structure is summarised in Fig. 6.1.

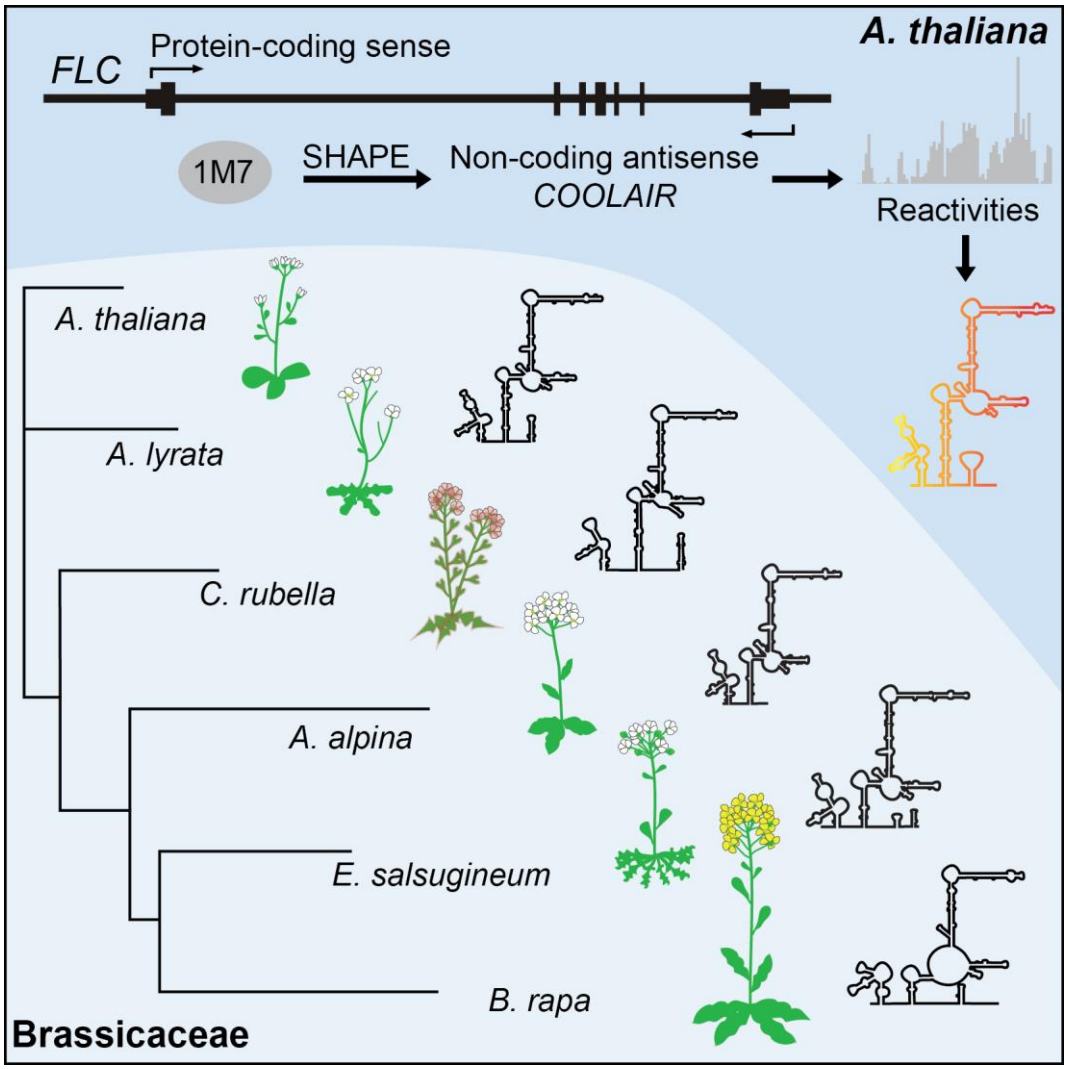


Figure 6.1: Evolutionary conservation of the distal *COOLAIR* secondary structure. Where, SHAPE refers to selective 2'-OH acetylation analyzed by primer extension, a method to chemically-probe the secondary structure of RNA molecules; 1M7 (1-methyl-7-nitroisatoic anhydride) is the reagent used in SHAPE chemical-probing; and the 'reactivities' profile represents the reactivity of each nucleotide in a given RNA sequence to 1M7. Phylogenetic tree as in Fig. 3.5. Adapted from Hawkes *et al.* (2016).

Conservation of structural elements in *HOTAIR* and *Braveheart* have been linked to functional roles (Somarowthu *et al.*, 2015; Xu *et al.*, 2016). Conserved structural domains may therefore play a role in the *COOLAIR* regulatory mechanism. In Chapter 5, we demonstrate conserved expression patterns with upregulation of *COOLAIR* transcripts during cold treatment (from *BrFLCA10* in transgenic plant lines). This supports a similar mechanism of *COOLAIR*-mediated downregulation of sense transcription during vernalization (Swiezewski *et al*., 2009). Functional conservation in the autonomous and vernalization pathways should be investigated further.

This works supports the functional significance of *COOLAIR*; however, it is still just one long non-coding RNA amongst thousands of uncharacterised sense and antisense transcripts in *A. thaliana* (Yamada *et al.*, 2003). What we have demonstrated is the importance of looking beyond primary nucleotide sequence in conservation analyses, and looking in greater detail at promoter regions, secondary structure, genetic architecture, and expression patterns. This research highlights the need to analyse all dimensions of conservation before disregarding ncRNAs as transcriptional noise (Diedrichs, 2014). The many novel mechanisms of lncRNA regulation thus far discovered, from interaction with single (Xue *et al.*, 2016) or multiple (Wierzbicki *et al.*, 2009; Tsai *et al.*, 2010) proteins to behaving as target mimics (Franco-Zorrilla *et al.*, 2007) or miRNA sponges (Du *et al.*, 2016), further support this recommendation for comprehensive analysis at an individual scale. Individual studies such as this should therefore not be viewed as archetypal, but can provide insight into methods for evaluating other lncRNAs and their mechanisms of regulation.

The best characterised lncRNAs are, for obvious reasons, those that are dysregulated in human cancers, heart and other diseases; for example, *HOTAIR* during breast cancer (Gupta *et al.*, 2010), *MALAT1* during lung cancer (Gutschner *et al.*, 2013A), and *H19* during colorectal cancer (Cui *et al.*, 2002) progression. Differential expression of lncRNAs has similarly been reported in plants experiencing abiotic or biotic stress (Amor *et al.*, 2009; Xin *et al.*, 2011; Wu *et al.*, 2012B; Shuai *et al.*, 2014; Wang *et al.*, 2014B; Zhang *et al.*, 2014; Zhu *et al.*, 2014). Continued study of lncRNAs in plants will therefore likely have important implications for plant breeding and food production. This work supports a role for *COOLAIR* in the vegetative-floral developmental transition across flowering plants (Chapter 3) and suggests upregulation in the cold is conserved (Chapter 5). This indicates ancient and universal roles for lncRNAs in regulating developmental transitions across both plant and animal kingdoms, and emphasises the necessity of crosstalk between the plant and animal research fields.

6.2 Searching for *COOLAIR* in flowering plants

At the start of this research project, and to the best of our knowledge, *COOLAIR* had not been detected from nor analysed in any species other than the reference plant, *A. thaliana*. In 2014, Castaings *et al.* detected *COOLAIR* in two other members of the Brassicaceae family: the proximal and distal forms from the *PEP1* homologue of *FLC* in *Arabis alpina*, and the proximal form from one of two *FLC* homologues in *A. lyrata*. They further suggested conservation of the promoter sequence in two other species, *C. rubella* and *A. halleri*. This study was important as it suggested that *COOLAIR* was not aberrant antisense transcription at the *FLC* locus but had been conserved over evolutionary time. This was supported by later work that identified *COOLAIR* transcription from one of four *FLC* homologues (*FLCA2*) in *B. rapa* (Li *et al.*, 2016), and from *C. rubella*, *A. lyrata* (including the distal transcript), *B. rapa* (three homologues), *B. oleracea* (three homologues), and *B. napus* (three of three tested homologues) (Chapters 3 and 4; Hawkes *et al.*, 2016). In contrast to the annual life habit of *A. thaliana*, *A. alpina* and *A. lyrata* are perennials and can flower multiple times over their life history. This reveals that *COOLAIR* has been conserved in plants with different life histories.

Together, these research findings suggest that *COOLAIR* was present at *FLC* in the original Brassicaceae ancestral palaeogenome, pre Camelineae/Calepineae split (Murat *et al.*, 2015). We cannot yet speculate on its origins before that or exactly when it arose in angiosperms. Antisense transcription may have arisen at the ancestral *FLC* gene after or at the same time as sense transcription (before would serve no purpose). *FLC* homologues have been identified in flowering dicots (Reeves *et al.*, 2007; Yi *et al.*, 2014; Kumar *et al.*, 2016; Taylor *et al.*, 2017) and even in some monocots (Ruelens *et al.*, 2013). Those interested in *COOLAIR* conservation in other families therefore have a rich set of *FLC*-like genes to choose from. The mechanism of downregulation in cold was reported to be conserved even where *FLC* function itself may have diversified (Reulens *et al.*, 2013; Taylor *et al.*, 2017). This is highly promising for *COOLAIR* conservation, given its role in early downregulation of the sense transcript during cold (Swiezewski *et al.*, 2009; Csorba *et al.*, 2014). Even where FLC protein function has diversified, we speculate that the in-built self-regulatory circuit of sense and antisense transcripts at *FLC*, described by Marquardt *et al.* (2014), would be retained.

COOLAIR transcripts are expressed at low levels under non-vernalized conditions, and we do not yet know if they are universally upregulated during cold. The technique developed in this thesis uses a series of nested touchdown PCRs to detect all splicing isoforms

without bias, even when expressed at very low levels in the genome (described in Chapter 2). This method was effectively used to detect *COOLAIR* distal transcripts from *A. lyrata*, and both proximal and distal transcripts from *FLCA10* and *FLCA3a* homologues in *B. rapa*, where other groups were unsuccessful (Castaings *et al.*, 2014; Li *et al.*, 2016). It is therefore recommended as an efficient and thorough means of checking for *COOLAIR* transcription in other species. In species where primary nucleotide conservation is low, we recommend combining this with secondary structure analysis to predict regions of syntenic transcription from which primers can be designed (Chapter 3; Hawkes *et al.*, 2016).

6.3 Potential mechanisms of *COOLAIR* regulation

Potential mechanisms for *COOLAIR* downregulation of *FLC* were discussed in Chapter 1. Experimental evidence strongly suggests that sense and antisense transcripts at *FLC* feedback to each other in a self-regulatory circuit (Liu *et al*., 2010; Marquardt *et al.*, 2014; Wang *et al.*, 2014A). Cotranscriptional coupling of RNA Pol II elongation rate, sense and antisense processing, and chromatin marks determine the expression state of *FLC* before cold. In the cold, *COOLAIR* promotes the removal of H3K36 active histone marks, helping to push *FLC* into a repressive state (Csorba *et al.*, 2014).

Vernalization reduces total *FLC* expression levels by switching cells off digitally; *FLC* is in either an on- or an off-state in any given cell (Angel *et al.*, 2015). In off-state cells, the locus is silenced and both sense and antisense transcripts are suppressed. In on-state cells, either antisense or sense are transcribed from each locus (Rosa *et al.*, 2016). In the cold, the number of on-state cells expressing antisense transcripts increases; this is followed by an increase in the number of off-state cells as cold progresses (Angel *et al.*, 2015; Rosa *et al.*, 2016). Transient expression of *PWR1* and *ICR1* lncRNAs are thought to initiate the transition to an on- or off-state, respectively, at the *FLO11* locus in yeast (Bumgarner *et al.*, 2012). Expression of *COOLAIR* may similarly increase the likelihood of a cell shifting to an off-state through cotranscriptional feedback mechanisms.

It has been assumed that the proximal transcript is a stronger repressor of sense transcription than the distal; increasing the ratio of proximal:distal transcripts correlates with sense downregulation before cold, and *COOLAIR* proximal upregulation is higher than distal in the cold (Swiezewski *et al.*, 2009; Liu *et al.*, 2010). It has been speculated that the distal transcript is unimportant for, a weak repressor of, or even an activator of sense transcription. The distal transcript folds into an elaborate and complex secondary structure (Hawkes *et al.*, 2016). It contains two right-hand turn (r-turn) motifs – an

asymmetric 5’ internal loop between two helices. Similar r-turns in other RNA species have been shown to act as receptors for proteins or to facilitate other interactions (Montemayor *et al.*, 2014; Ren at al., 2016). In Chapter 3 and in Hawkes *et al.* (2016), we revealed conservation of the global secondary structure of the distal transcript, including the two r-turns, supporting a functional role.

SHAPE-probing experiments did not detect differences in distal or proximal secondary structures in the cold, therefore we do not think *COOLAIR* is acting as a thermosensor in this manner (Karissa Sanbomatsu, unpublished). One hypothesis is that the distal r-turns facilitates binding with regulatory proteins or chromatin remodellers, such as a H3K36 demethylase. Potentially, loss of the r-turn (without the extra exon) or change in its position (with the extra exon) in the Var2-6 ecotype inhibits this interaction, contributing to its late flowering phenotype (Li *et al.*, 2015; Hawkes *et al.*, 2016). In Chapter 4, we link the length of helix H4, which supports the first t-rurn, with differences in expression of the *Brassica FLC* homologues. Association between a short or non-existent helix H4 and high *FLC* sense expression (aka. weak *COOLAIR* regulation) is supported by the analysis of *FLCA10* transgenic lines with splice site mutations (Chapter 5). Large ‘clouds’ of antisense transcripts accumulate at and remain tethered to the *FLC* locus during cold (Rosa *et al.*, 2016). Another possibility then, is that *COOLAIR* structure physically disrupts sense transcription, perhaps by blocking access to RNA Pol II or a H3K36 methyltransferase. Highly structured regions of RNA were found to correlate with lower translation efficiency in 5’ regions and alternate splicing and polyadenylation *in vivo* in *A. thaliana* (Ding *et al.*, 2014).

H3K36 demethylation in the nucleation region of *FLC* is important for vernalization-mediated repression, but we do not know if it plays a role in setting initial levels of *FLC*. It may be that the distal transcript regulates *FLC* via a different mechanism in the warm. Alternate mechanisms are possible because of cold-specific changes at *FLC*: *COOLAIR* expression is upregulated and unspliced transcripts accumulate at the locus (Swiezewski *et al.*, 2009; Rosa *et al.*, 2016), the 5’ to 3’ gene loop is lost (Crevillén *et al.*, 2013), and a new 5’ intron 1 gene loop forms (Kim and Sung, 2017). Li *et al.* (2015) associated altered distal processing in Var2-6 with increased *FRI*-mediated 5’ capping of the sense transcript under non-vernalized conditions. Assuming the distal is a negative regulator of the sense, this suggests either that the altered distal spliceoform promotes 5’ capping, or that the canonical spliceoform inhibits it. This could also involve secondary structure, with loss of the highly structured first r-turn in the Var2-6/*BrFLCA10* alleles eliminating obstruction of

5’ capping, for example. It is further possible the same mechanism is at work before and after vernalization, but that cold-associated changes at *FLC* amplify its effect. Further research should focus on identifying the H3K36 demethylase at *FLC*, testing for *COOLAIR*-protein interactions, and elucidating the distal *in vivo* secondary structure.

Chapter 4 revealed that *COOLAIR* is transcribed from multiple *FLC* homologues within *Brassica*. *COOLAIR* was undetectable from *FLC* homologues and/or accessions with low sense expression. This parallels the correlation between sense and antisense transcription observed in *Arabidopsis* (Swiezewski *et al.*, 2009). *COOLAIR* transcripts from *BrFLCA10* (Chapter 5) and *BrFLCA2* (Li *et al.*, 2016) are upregulated in the cold; we speculate that this cold-response has been conserved across homologues. *BrFLCA10* and the Var2-6 *FLC* allele share a non-canonical distal splice site; we show in Chapter 5 that they also exhibit similar expression patterns and behaviour. Together, these results support similar mechanisms of regulation by *COOLAIR* in *Brassica* and *Arabidopsis*, despite their evolutionary distance of ~43 Ma.

Further work is required to determine whether *COOLAIR* regulates its own locus in *cis*, or other loci in *trans*. The data thus far, in both *Brassica* and *Arabidopsis*, points to the former. Different *COOLAIR* isoforms have evolved at distinct *FLC* loci; we hypothesise that expression at each will reflect this. This is supported by circumstantial evidence (correlation between different *COOLAIR* isoforms and *FLC* expression), and by transgenic experiments in *FLCA10*. It is further supported by work in *Arabidopsis*: evidence for self-regulatory feedback between sense and antisense and smFISH experiments, where *COOLAIR* is found specifically at the locus from which it was transcribed and not elsewhere in the nucleus (Marquardt *et al.*, 2014; Wang *et al.*, 2014A; Rosa *et al.*, 2016). *Brassica FLC* homologues without antisense transcription (like the *Arabidopsis* TEX lines described by Csorba *et al.*, 2014) could be generated to test this. Differences in *FLC* sense expression and flowering time with and without antisense transcription could then be compared within and between homologues. The generated *BoFLCC2* and *BrFLCA3a* transgenic lines should be further analysed at T3 to check for small differences in sense expression.

Li *et al.* (2016) were only able to detect *COOLAIR* from *FLCA2* in *B. rapa*; they hypothesised that it regulates *BrFLCA2* in *cis* and the other homologues in *trans*. The possibility of *trans*-regulation by a single dominant *COOLAIR* species is an intriguing one, but is not supported by the existence of *COOLAIR* at other homologues and our

understanding of sense-antisense feedback. We would therefore modify this to suggest that each *COOLAIR* regulates its own copy in *cis*, potentially in addition to *trans*-regulation from a dominant species. We predict that *COOLAIR* from *BrFLCA2* is a strong negative regulator from secondary structure analysis. This supports the hypothesis that the *BrFLCA2* copy is dominant (Li *et al.*, 2016). Overexpression of *BrFLCA2* antisense transcripts in *B. rapa* pekinensis, a vernalization-requiring biennial, downregulated expression from the other *FLC* homologues and promoted flowering (Li *et al*., 2016). Although this suggests *trans*-regulation by *COOLAIR* is possible in an overexpression system, it does not establish whether it occurs within the wildtype *Brassica* genome.

Trans-regulation of *COOLAIR* may have evolved specifically in species with duplicated genomes and multiple copies of *FLC*. Extensive genome reshuffling after genome triplication and subsequent species diversification in *Brassica* likely facilitated the development of new mechanistic pathways (Osborn *et al.*, 2003; Gaeta *et al*., 2007; Cheng *et al.*, 2013). *Trans*-regulation may further be a consequence of subgenome dominance i.e. of one of the three ancestral (n=7) subgenomes (Wang *et al.*, 2011). In the more recently duplicated *B. napus* species, it is interesting to speculate that *trans*-regulation is more likely between the *FLC* homeologues because of their high sequence identities (Gaeta *et al*., 2007). *Trans*-regulation of *FLC* homeologues was supported by A versus C genome expression patterns in the cold. Expression analysis of *FLC* copies after removal or suppression of their homeologues would help to validate this.

6.4 An evolutionary tool to finetune flowering time

This thesis demonstrates that natural variation in *COOLAIR* has arisen multiple times over evolutionary history, in both *Brassica* and *Arabidopsis*. Thus far, the flowering phenotype of only one *Arabidopsis FLC* haplotype group has been confirmed to work through *COOLAIR* (Li *et al.*, 2015), although others may still be identified. *Brassica* species provide an extensive system in which to analyse variation at *COOLAIR*; large breeding populations of *Brassica* cultivars display diverse flowering phenotypes, each with multiple copies of *FLC* (Schiessl *et al*., 2014; Schiessl *et al*., 2017). In Chapter 5, several SNPs with the potential to alter *COOLAIR* splicing or polyadenylation were identified from three *FLC* genes in a subset of 32 winter and 13 spring *B. napus* cultivars.

We propose that *COOLAIR* could act as a regulatory tool to finetune flowering time. Long non-coding RNAs by definition do not encode proteins and therefore are often less sensitive to deleterious mutations. Nucleotide substitution rates are higher in ncRNAs than

protein-coding transcripts; they therefore provide a platform for rapid adaptation to changes in the environment (Pang *et al.*, 2006; Kutter *et al.*, 2012; Nitsche *et al.*, 2015). This same trait gives lncRNAs a more transitionary nature. High rates of transcriptional turnover in rodents is thought to have driven the loss of nearly 50% of intergenic lncRNA loci since the last common rat-mouse ancestor (Kutter *et al.*, 2012). Rapid evolution at lncRNAs in polyploid species, such as *Brassica*, would be accelerated by the presence of multiple gene copies. Diverse functions have, for example, arisen in lncRNAs from *roX1* paralogues after gene duplication in *Drosophila* (Quinn *et al.*, 2016).

Gradations in *COOLAIR* regulatory strength has been hypothesised; for example, through changes in the length of helix H4 strengthening or weakening its activity. This could facilitate more subtle adaptations to flowering time. Even minor variations could provide a selective advantage in a large population, ensuring seed set and dispersal at optimum conditions. For example, differences in optimum vernalization temperature between Swedish *A. thaliana* accessions helps to ensure cold-saturation is reached before snowfall (Duncan *et al*., 2015). Experiments were set up in Chapter 5 to modify *COOLAIR* in *BoFLCC2*, *BrFLCA3a* and *BrFLCA10* through one or a low number of *cis* non-coding polymorphisms. These were only effective in *BrFLCA10* transgenic lines, where the mutations were sufficient to bring about significant changes in *COOLAIR* structure through altered splicing. An alternate hypothesis then is that multiple mutations (or localised mutations in a functionally-critical region) accumulate over time to bring about effective phenotypic change.

We have identified a splicing mutation that has arisen independently in both *Brassica* (the *FLCA10* homologue) and *Arabidopsis* (Var2-6) associated with late flowering. Inhibiting formation of this spliceoform switches the plant from a late to early flowering phenotype. This is exciting not only because it indicates similar mechanisms of *COOLAIR* regulation across species, but because it pinpoints this region (H4, the first r-turn and H5) as critical for *COOLAIR* function. The work presented here and in Li *et al.* (2015) confirm that flowering time adaptations can work through *COOLAIR*. It is not yet known how frequently this occurs, or how easy it is. To help answer the former, the region of interest should be sequenced in a large number of other species and cultivars. To answer the latter, further analysis of polymorphisms at other *COOLAIR* regions is required.

6.5 A biotechnological tool to finetune flowering time

This research sheds light on *FLC* dynamics in *Brassica* flowering time. In agreement with a recent publication from Schiessl *et al.* (2017), we suggest that *FLCA10* is important for the spring/winter divide and setting non-vernalized flowering time in *Brassica napus*. We correlate high expression of this homologue in winter lines with a weak *COOLAIR* regulator, and low expression in spring lines with a non-functional protein. We further propose that the C genome *FLC* copies, *BnFLCC2* and *BnFLCC3a*, are critical for measuring cold length during vernalization. These reactivated after a short (two week) period of cold, whereas A genome copies were silenced. A role for *FLCC2* in vernalization response is supported by studies in *B. oleracea* (Ridge *et al.*, 2015; Irwin *et al.*, 2016). Subgenome dominance and potential mechanisms of homeologue repression in *Brassica napus* will prove exciting for future investigations. This research further sheds light on the role of *COOLAIR* in *Brassica FLC* regulation and the extra level of complexity it brings. An appreciation of *COOLAIR* is necessary for understanding the role each *FLC* homologue plays in contributing to absolute flowering time.

Consistency and uniformity of flowering time or heading date, high yield, and climatic robustness are traits that are selectively bred into *Brassica* vegetables and oilseeds (Jung and Müller, 2009). Whereas transcriptional activators were targeted in the domestication of crops, protein-coding genes have primarily been targeted in crop diversification (Alonso-Blanco *et al.*, 2009). Marker assisted breeding at *FLC* has been used to select for late or early flowering plant varieties (Bayer CropScience, personal communication). Identification of *COOLAIR*-coding regions unlocks additional layers of variation to select from, in addition to helping prevent the introduction of non-desired effects. *COOLAIR* responds to and is upregulated by cold; vernalization can occur at a range of temperatures, but not above 14 °C in *Arabidopsis* and optimum temperatures are often ecotype-specific (Duncan *et al.*, 2015). Milder winters are predicted in mid- and high-latitude countries in consequence of anthropogenic climate change over the next hundred years (IPCC, 2013). This may have consequences for *COOLAIR* upregulation and the vernalization-mediated silencing of *FLC*. Indeed, rises in global temperature are already thought to have impacted natural flowering populations (Fitter and Fitter, 2002; Cook *et al*., 2012; Rafferty and Nabity, 2017). It is therefore important to develop varieties that can overwinter at higher temperatures for consistent flowering.

If *COOLAIR* is as an evolutionary tool to adapt flowering time to changing climate, it could potentially serve a similar role in the artificial generation of crop varieties with

different flowering times. Desired changes in flowering time could be introduced through mutations at *COOLAIR* or, as demonstrated by Li *et al.* (2016), through overexpression lines. Is this, however, any more useful than directly manipulating the FLC protein itself? It is possible that, as discussed above, more subtle variations in flowering time could be manufactured through *COOLAIR*. Engineering weaker *COOLAIR* regulators could facilitate the development of late flowering lines, and stronger *COOLAIR* regulators, early flowering lines, as demonstrated through the *BrFLCA10* transgenics developed in Chapter 5. Genetic modification or marker-assisted breeding could be used to generate flowering time variation through targeting *COOLAIR*. A difficulty with *Agrobacterium*-mediated transformation systems in plants, such as that demonstrated in this thesis, is that there is no control over where the transgene is inserted. More precise genome editing technologies have been developed to directly modify endogenous gene targets; these include zinc finger nucleases (ZFNs), TAL effector nucleases (TALENs), and most recently, CRISPR-Cas9 technology (Doudna and Charpentier, 2014). The CRISPR-associated Cas9 endonuclease introduces a double-stranded break in target DNA *in vivo* via site-specific binding of an RNA guide sequence; the guide sequence can be modified to target any sequence of interest (Doudna and Charpentier, 2014). CRISPR-Cas9 has been successfully used to introduce homozygous multiplex mutations in dicot and monocot plants; it is therefore a promising technology for the development of new lines with desired agronomic traits (Ma *et al.*, 2015).

Development of a method that allows direct control of flowering time through, for example, topical applications of *COOLAIR* to promote or delay flowering, would be more practical for responding quickly to unforeseen changes in annual climate. Topical RNAi technologies are being developed, i.e. BioDirect™ from Monsanto, but the uptake of topically-applied RNA into plant cells has not been proven. Even if this technology were to exist, the highly-interconnected nature of sense and antisense regulation at *FLC* suggests that topically applied or injected *COOLAIR* lncRNAs would be less (or not at all) effective in regulating *FLC* chromatin; nonetheless such approaches are worthy of investigation.

6.6 Conclusion

We propose that conserved *COOLAIR* transcripts in *Brassica* have evolved into distinct regulators of their *FLC* homologues over millions of years, helping to maintain unique sense expression and associated protein levels at that locus. Allele-specific changes in *COOLAIR* structure and processing have arisen over the last 10 thousand years in *B. napus*, and may drive phenotypic differences in flowering time. *COOLAIR* long non-coding RNAs may therefore act as a novel tool to finetune flowering time; this could be adapted for the selection of agronomic traits.

Future research should focus on: (1) elucidating the role of structure in the distal *COOLAIR* function, (2) confirming the relationship between *COOLAIR* processing and flowering time in *Brassica* species, and (3) analysis of the frequency of phenotypically significant mutations at *COOLAIR* in *Brassica* and other species. The above studies would further understanding of how the distal *COOLAIR* transcript functions in *Arabidopsis* and *Brassica*, and determine how useful *COOLAIR* is for manipulating flowering time.

7 References

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., Araki, T. 2005. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309: 1052-1056

Ågren, J., Oakley, C.G., Lundemo, S. and Schemske, D.W. 2017. Adaptive divergence in flowering time among natural populations of *Arabidopsis thaliana*: Estimates of selection and QTL mapping. Evolution 71: 550-564

Albani, M.C., Castaings, L., Wötzel, S., Mateos, J.L., Wunder, J., Wang, R., Reymond, M., Coupland, G. 2012. PEP1 of *Arabis alpina* is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. PLoS Genetics 8: e1003130

Allen, E., Xie, Z., Gustafson, A.M., Carrington, J.C. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. Cell 121: 207-221

Alló, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A., Klinck, R., Chabot, B., Kornblihtt, A.R. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nature Structural and Molecular Biology 16: 717-724

Alonso-Blanco, C., Aarts, M.G., Bentsink, L., Keurentjes, J.J., Reymond, M., Vreugdenhil, D., Koornneef, M. 2009. What has natural variation taught us about plant development, physiology, and adaptation? The Plant Cell 21: 1877-1896

Amasino, R., 2010. Seasonal and developmental timing of flowering. The Plant Journal 61: 1001-1013

Amor, B.B., Wirth, S., Merchan, F., Laporte, P., d'Aubenton-Carafa, Y., Hirsch, J., Maizel, A., Mallory, A., Lucas, A., Deragon, J.M., Vaucheret, H. 2009. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Research 19: 57-69

Andersson, C.R., Helliwell, C.A., Bagnall, D.J., Hughes, T.P., Finnegan, E.J., Peacock, W.J., Dennis, E.S. 2008. The *FLX* gene of *Arabidopsis* is required for *FRI*-dependent activation of *FLC* expression. Plant and Cell Physiology 49: 1910299

Andrés, F., Porri, A., Torti, S., Mateos, J., Romera-Branchat, M., García-Martínez, J.L., Fornara, F., Gregis, V., Kater, M.M., Coupland, G. 2014. SHORT VEGETATIVE PHASE reduces gibberellin biosynthesis at the *Arabidopsis* shoot apex to regulate the floral transition. PNAS 111: 2760-2769

Andrés, F. and Coupland, G. 2012. The genetic basis of flowering responses to seasonal cues. Nature Reviews Genetics 13: 627-639

Angel, A., Song, J., Dean, C., Howard, M. 2011. A Polycomb-based switch underlying quantitative epigenetic memory. Nature 476: 105-108

Angel, A., Song, J., Yang, H., Qüesta, J.I., Dean, C., Howard, M. 2015. Vernalizing cold is registered digitally at *FLC*. PNAS 112: 4146-4151

Ausín, I., Alsonso-Blanco, C., Jarillo, J.A., Ruiz-García, L., Martínez-Zapater, J.M. 2004. Regulation of flowering time by FVE, a retinoblastoma-associated protein. Nature Genetics 36: 162-166

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., Dean, C. 2004. Vernalization requires epigenetic silencing of *FLC* by histone methylation. Nature 427: 165-167

Bäurle, I., and Dean, C. 2008. Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of *Arabidopsis* targets. PLoS ONE 3: e2733

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., Mathews, S. 2010. Dated molecular phylogenies indicated a Miocene origin for *A. thaliana*. PNAS 107: 18724-18728

Ben-Shem, A., Jenner, L., Yusupova, G., Yusupov, M. 2010. Crystal structure of the eukaryotic ribosome. Science 330: 1203-1209

Bentley, D.L. 2014. Coupling mRNA processing with transcription in time and space. Nature Reviews Genetics 15: 163-175

Berry, S., Hartley, M., Olsson, T.S., Dean, C., Howard, M. 2015. Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance. eLIFE 4: e07205

Beyer, A.L. and Osheim, Y.N. 1988. Splice site selection, rate of splicing and alternative splicing on nascent transcripts. Genes & Development 2: 754-765

Bezerra, I.C., Michaels, S.D., Schomburg, F.M., Amasino, R.M. 2004. Lesions in the mRNA cap-binding gene *ABA HYPERSENSITIVE 1* suppress *FRIGIDA*-mediated delayed flowering in *Arabidopsis*. The Plant Journal 40: 112-119

Blázquez, M.A., Ahn, J.H., Weigel, D. 2003. A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. Nature Genetics 33: 168-171

Blázquez, M.A., Soowal, L.N., Lee, I., Weigel, D. 1997. *LEAFY* expression and flower initiation in *Arabidopsis*. Development 124: 3835-3844

Bohuon, E.J.R., Ramsay, L.D., Craft, J.A., Arthur, A.E., Marshall, D.F., Lydiate, D.J., Kearsey, M.J. 1998. The association of flowering time quantitative trait loci with duplicated regions and candidate loci *in Brassica oleracea*. Genetics 150: 393-491

Bomblies, K. and Madlung, A. 2014. Polyploidy in the *Arabidopsis* genus. Chromosome Research 22: 117-134

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 123: 1279-1291

Bowers, J.E., Chapman, B.A., Rong, J., Paterson, A.H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433-438

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitzm E.M., Smyth, D.R. 1993. Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development 119: 721-743

Box, M.S., Coustham, V., Dean, C., Mylne, J.S. 2011. Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from *Arabidopsis*. Plant Methods 7: 7

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., Coen, E. 1997. Inflorescence commitment and architecture in *Arabidopsis*. Science 275: 80-83

Brodersen, D.E., Clemons, W.M., Carter, A.P., Wimberly, B.T., Ramakrishnan, V. 2002. Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. Journal of Molecular Biology 316: 725-768

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green E.D., Sidow, A., Batzoglou, S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Research 13: 721-731

Bumgarner, S.L., Dowell, R.D., Grisafi, P., Gifford, D.K., Fink, G.R. 2009. Toggle involving *cis*-interfering noncoding RNAs controls variegated gene expression in yeast. PNAS 106: 18321-18326

Bumgarner, S.L., Neuert, G., Voight, B.F., Symbor-Nagrabska, A., Grisafi, P., van Oudenaarden, A., Fink, G.R. 2012. Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. Molecular Cell 45: 470-482

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., Neidle, S. 2006. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Research 34: 5402-5415

Burn, J.E., Bagnell, D.J., Metzger, J.D., Dennis, E.S., Peacock, W.J. 1993. DNA methylation, vernalization, and the initiation of flowering. PNAS 90: 287-291

Cáceres, J.F. and Kornblihtt, A.R. 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends in Genetics 18: 186-193

Caicedo, A.L., Stinchcombe, J.R., Olsen, K.M., Schmitt, J., Purugganan, M.D. 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. PNAS 101: 15670-15675

Campalans, A., Kondorosi, A., Crespi, M. 2004. *Enod40*, a short open reading frame–containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. The Plant Cell 16: 1047-1059

Carruthers, J.M., Cook, S.M., Wright, G.A., Osborne, J.L., Clark, S.J., Swain, J.L., Haughton, A.J. 2017. Oilseed rape (*Brassica napus*) as a resource for farmland insect pollinators: quantifying floral traits in conventional varieties and breeding systems. GCB Bioenergy 9: 1370-1379

Castaings, L., Bergonzi, S., Albani, M.C., Kemi, U., Savolainen, O., Coupland, G. 2014. Evolutionary conservation of cold-induced antisense RNAs of *FLOWERING LOCUS C* in *Arabidopsis thaliana* perennial relatives. Nature Communications 5: 4457

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345: 950-953

Chandler, J., Martinez-Zapater, J.M., Dean, C. 1999. Mutations causing defects in the biosynthesis and response to gibberellins, abscisic acid and phytochrome B do not inhibit vernalization in *Arabidopsis fca-1*. Planta 210: 677-682

Chandler, J., Wilson, A., Dean, C. 1996. *Arabidopsis* mutants showing an altered response to vernalization. The Plant Journal 10: 637-644

Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., Guttman, M. 2016. *Xist* recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science 354: 468-472

Chen, L., Cheng, J-C., Castle, L., Sung, Z.R. 1997. *EMF* genes regulate *Arabidopsis* inflorescence development. The Plant Cell 9: 2011-2024

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M.A., Wang, X. 2013. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. The Plant Cell 25: 1541-1554

Cheng, F., Wu, J., Wang, X. 2014. Genome triplication drove the diversification of *Brassica* plants. Horticulture Research 1: 14024

Chi, K.R. 2016. The dark side of the human genome. Nature 538: 275-277

Choi, K., Kim, J., Hwang, H-J., Kim, S., Park, C., Kim, S.Y., Lee, I. 2011. The FRIGIDA complex activates transcription of *FLC*, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. The Plant Cell 23: 289-303

Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A., Rozowsky, J.S. 2011. The reality of pervasive transcription. PLoS Biology 9: e1000625

Clarke, J.H. and Dean, C. 1993. Mapping *FRI*, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. Molecular and General Genetics 242: 81-89

Clough, S.J. and Bent, A.F. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal 16: 735-743

Cook, B.I., Wolkovich, E.M., Parmesan, C. 2012. Divergent responses to spring and winter warming drive community level flowering trends. PNAS 109: 9000-9005

Corley, M., Solem, A., Qu, K., Chang, H.Y., Laederach, A. 2015. Detecting riboSNitches with RNA folding algorithms: a genome-wide benchmark. Nucleic Acids Research 43: 1859-1868

Coustham, V., Li, P., Strange, A., Lister, C., Song, J., Dean, C. 2012. Quantitative modulation of polycomb silencing underlies natural variation in vernalization. Science 337: 584-587

Crevillén, P., Sonmez, C., Wu, Z., Dean, C. 2013. A gene loop containing the floral repressor *FLC* is disrupted in the early phase of vernalization. The EMBO Journal 32: 140-148

Csorba, T., Questa, J.I., Sun, Q., Dean, C. 2014. Antisense *COOLAIR* mediates the coordinated switching of chromatin states at *FLC* during vernalization. PNAS 111:16160-16165

Cubas, P., Vincent, C., Coen, E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157-161

Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C.L., Feinberg, A.P. 2002. Loss of imprinting in colorectal cancer linked to hypomethylation of *H19* and *IGF2*. Cancer Research 62: 6442-6446

da Rocha, S.T. and Heard, E. 2017. Novel players in X inactivation: insights into *Xist*-mediated gene silencing and chromosome conformation. Nature Structural & Molecular Biology 24: 197-204

Davidovich, C., Zheng, L., Goodrich, K.J., Cech, T.R. 2013. Promiscuous RNA binding by Polycomb repressive complex 2. Nature Structural & Molecular Biology 20: 1250-1257

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., Kornblihtt, A.R. 2003. A slow RNA polymerase II affects alternative splicing *in vivo*. Molecular Cell 12: 525-532

De Lucia, F., Crevillen, P., Jones, A.M., Greb, T., Dean, C. 2008. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of *FLC* during vernalization. PNAS 105: 16831-16836

D'hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., Da Silva, C. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488: 213-217

Diederichs, S. 2014. The four dimensions of noncoding RNA conservation. Trends in Genetics 30: 121-123

Digel, B., Pankin, A., von Korff, M. 2015. Global transcriptome profiling of developing leaf and shoot apices reveals distinct genetic and environmental control of floral transition and inflorescence development in barley. The Plant Cell 27: 2318-2334

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., Assmann, S.M. 2014. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505: 696-700

Doudna, J.A. and Charpentier, E. 2014. The new frontier of genome engineering with CRISPR-Cas9. Science 346: 1258096

Doyle, M.R., Bizzell, C.M., Keller, M.R., Michaels, S.D., Song, J., Noh, Y.S., Amasino, R.M. 2005. *HUA2* is required for the expression of floral repressors in *Arabidopsis thaliana*. The Plant Journal 41: 376-385

Du, Z., Sun, T., Hacisuleyman, E., Fei, T., Wang, X., Brown, M., Rinn, J.L., Lee, M.G.S., Chen, Y., Kantoff, P.W., Liu, X.S. 2016. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. Nature Communications 7: 10982

Duncan, S., Holm, S., Questa, J., Irwin, J., Grant, A., Dean, C. 2015. Seasonal shift in timing of vernalization as an adaptation to extreme winter. eLIFE 4: e06620

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., Maizels, N. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes & Development 18: 1618-1629

Eamens, A., Wang, M.B., Smith, N.A., Waterhouse, P.M. 2008. RNA silencing in plants: yesterday, today, and tomorrow. Plant Physiology 147: 456-468

Edwards, K., Johnstone, C., Thompson, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Research 19: 1349

Eißmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Groß, M., Schirmacher, P., Rippe, K., Braun, T., Zörnig, M., Diederichs, S. 2012. Loss of the abundant nuclear non-coding RNA *MALAT1* is compatible with life and development. RNA Biology 9: 1076-1087

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74

Eriksson, S., Böhlenius, H., Moritz, T., Nilsson, O. 2006. GA4 is the active gibberellin in the regulation of *LEAFY* transcription and *Arabidopsis* floral initiation. The Plant Cell 18: 2172-2181

Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., Tress, M.L. 2014. Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. Human Molecular Genetics 23: 5866-5878

Finnegan, E.J., Sheldon, C.C., Jardinaud, F., Peacock, W.J., Dennis, E.S. 2004. A cluster of *Arabidopsis* genes with a coordinate response to an environmental stimulus. Current Biology 14: 911-916

Fitter, A.H. and Fitter, R.S.R. 2002. Rapid changes in flowering time in British plants. Science 296: 1689-1691

Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N., Banfield, M.J., Rameau, C. 2003. *DETERMINATE* and *LATE FLOWERING* are two *TERMINAL FLOWER1*/*CENTRORADIALIS* homologs that control two distinct phases of flowering initiation and development in pea. The Plant Cell 15: 2742-2754

Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G., Putterill, J. 1999. *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. The EMBO Journal 18: 4679-4688

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., Paz-Ares, J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics 39: 1033-1037

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Research 32: 273-279

Gaeta, R.T., Pires, J.C., Iniguez-Luy, F., Leon, E., Osborn, T.C. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. The Plant Cell 19: 3403-3417

Galupa, R. and Heard, E. 2015. X-chromosome inactivation: new insights into *cis* and *trans* regulation. Current Opinion in Genetics & Development 31: 57-66

Gan, E.S., Xu, Y., Wong, J.Y., Goh, J.G., Sun, B., Wee, W.Y., Huang, J., Ito, T. 2014. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of *FLC* in *Arabidopsis*. Nature Communications 5: 5098

Gautheret, D.A., Konings, D.A., Gutell, R.R. 1995. GU base pairing motifs in ribosomal RNA. RNA 1: 807-814

Gazzani, S., Gendall, A.R., Lister, C. 2003. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiology 132: 1107-1114

Gendall, A.R., Levy, Y.Y., Wilson, A., Dean, C. 2001. The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. Cell 107: 525-535

Geraldo, N., Bäule, I., Kidou, S-i., Hu, X., Dean, C. 2009. FRIGIDA delays flowering in *Arabidopsis* via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex. Plant Physiology 150: 1611-1618

Ginzinger, D.G. 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Experimental Hematology 30: 503-512

Gosai, S.J., Foley, S.W., Wang, D., Silverman, I.M., Selamoglu, N., Nelson, A.D., Beilstein, M.A., Daldal, F., Deal, R.B., Gregory, B.D. 2015. Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. Molecular Cell 57: 376-388

Graur, D., Zheng, Y., Price, N., Azevedo, R.B., Zufall, R.A., Elhaik, E. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biology and Evolution 5: 578-590

Greb, T., Mylne, J.S., Crevillén, P., Geraldo, N., An, H., Gendall, A.R., Dean, C. 2007. The PHD finger protein VRN5 functions in the epigenetic silencing of *Arabidopsis FLC*. Current Biology 17: 73-78

Greenup, A., Peacock, W.J., Dennis, E.S., Trevaskis, B. 2009. The molecular biology of seasonal flowering-responses in *Arabidopsis* and the cereals. Annals of Botany 103: 1165-1172

Gregory, R.I., Chendrimada, T.P., Cooch, N., Shiekhattar, R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell 123: 631-640

Gu, X., Jiang, D., Yang, W., Jacob, Y., Michaels, S.D., He, Y 2011. *Arabidopsis* homologs of retinoblastoma-associated protein 46/48 associate with a histone deacetylase to act redundantly in chromatin silencing. PLoS Genetics 7: e1002366.

Guo, Y.L., Todesco, M., Hagmann, J., Das, S., Weigel, D. 2012. Independent *FLC* mutations as causes of flowering time variation in *Arabidopsis thaliana* and *Capsella rubella*. Genetics 192: 729-739

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M-C., Hung, T., Argani, P., Rinn, J.L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R.B., Van de Vijver, M.J., Sakumar, S., Chang, H.Y. 2010. Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. Nature 464: 1071-1076

Gutschner, T., Hämmerle, M., Diederichs, S. 2013**A**. *MALAT1*—a paradigm for long noncoding RNA function in cancer. Journal of Molecular Medicine 91: 791-801

Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Groß, M., Zörnig, M. 2013**B**. The noncoding RNA *MALAT1* is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Research 73: 1180-1189

Haddad, F., Qin, A.X., Giger, J.M., Guo, H., Baldwin, K.M. 2007. Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. BMC Biotechnology 7: 21

Harris, K.A., Lünse, C.E., Li, S., Brewer, K.I., Breaker, R.R. 2015. Biochemical analysis of pistol self-cleaving ribozymes. RNA 21: 1852-1858

Hawkes, E.J., Hennelly, S.P., Novikova, I.V., Irwin, J.A., Dean, C., Sanbonmatsu, K.Y. 2016. *COOLAIR* antisense RNAs form evolutionarily conserved elaborate secondary structures. Cell Reports 16: 3087-3096

He, Y., Doyle, M.R., Amasino, R.M. 2004. PAF1-complex-mediated histone methylation of *FLOWERING LOCUS C* chromatin is required for the vernalization-response, winter-annual habit in *Arabidopsis*. Genes & Development 18: 2774-2784

He, Y., Michaels, S.D., Amasino, R.M. 2003. Regulation of flowering time by histone acetylation in *Arabidopsis*. Science 302: 1751-1754

Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouzé, P., Brunak, S. 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Research 24: 3439-3452

Hecht, V., Foucher, F., Ferrándiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Beltrán, J.P., Rameau, C., Weller, J.L. 2005. Conservation of *Arabidopsis* flowering genes in model legumes. Plant Physiology 137: 1420-1434

Helliwell, C.A., Robertson, M., Finnegan, E.J., Buzas, D.M., Dennis, E.S. 2011. Vernalization-repression of *Arabidopsis FLC* requires promoter sequences but not antisense transcripts. PLoS ONE 6: e21513

Helliwell, C.A., Wood, C.C., Robertson, M., Peacock, J.W., Dennis, E.S. 2006. The *Arabidopsis* FLC protein interacts directly *in vivo* with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. The Plant Journal 46: 183-192

Hepworth, J. and Dean, C. 2015. *Flowering Locus C*'s lessons: conserved chromatin switches underpinning developmental timing and adaptation. Plant Physiology 168: 1237-1245

Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A., Coupland, G. 2002. Antagonistic regulation of flowering-time gene *SOC1* by CONSTANS and FLC via separate promoter motifs. The EMBO Journal 21: 4327-4337

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., Ohta, K. 2008. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature 456: 130-134

Hirschi, A., Martin, W.J., Luka, Z., Loukachevitch, L.V., Reiter, N.J. 2016. G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. RNA 22: 1250-1260

Hornyik, C., Terzi, L.C., Simpson, G.G. 2010. The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. Developmental Cell 18: 203-213

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

Huang, T., Böhlenius, H., Eriksson, S., Parcy, F., Nilsson, O. 2005. The mRNA of the *Arabidopsis* gene *FT* moves from leaf to shoot apex and induces flowering. Science 309: 1694-1696

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., Attardi, L.D. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 142: 409-419

Huet, T., Miannay, F.A., Patton, J.R., Thore, S. 2014. Steroid receptor RNA activator (SRA) modification by the human pseudouridine synthase 1 (hPus1p): RNA binding, activity, and atomic model. PloS ONE 9: e94610

Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P., Brosius, J. 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. The EMBO Journal 20: 2943-2953

Imaizumi, T., Tran, H.G., Swartz, T.E., Briggs, W.R., Kay, S.A. 2003. FKF1 is essential for photoperiodic-specific light signalling in *Arabidopsis*. Nature 426: 302-306

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Ietswaart, R., Rosa, S., Wu, Z., Dean, C., Howard, M. 2017. Cell-size-dependent transcription of *FLC* and its antisense long non-coding RNA *COOLAIR* explain cell-to-cell expression variation. Cell Systems 4: 622-635

Irimia, M., Penny, D., Roy, S.W. 2007. Coevolution of genomic intron number and splice sites. Trends in Genetics 23: 321-325

Irwin, J.A., Lister, C., Soumpourou, E., Zhang, Y., Howell, E.C., Teakle, G., Dean, C. 2012. Functional alleles of the flowering time regulator *FRIGIDA* in the *Brassica oleracea* genome. BMC Plant Biology 12: 21

Irwin, J.A., Soumpourou, E., Lister, C., Ligthart, J.D., Kennedy, S., Dean, C. 2016. Nucleotide polymorphism affecting *FLC* expression underpins heading date variation in horticultural *Brassica*s. The Plant Journal 87: 597-605

Jaeger, K.E., Pullen, N., Lamzin, S., Morris, R.J., Wigge, P.A. 2013. Interlocking feedback loops govern the dynamic behavior of the floral transition in *Arabidopsis*. The Plant Cell 25: 820-833

Jodoin, R., Bauer, L., Garant, J.M., Mahdi Laaref, A., Phaneuf, F., Perreault, J.P. 2014. The folding of 5'-UTR human G-quadruplexes possessing a long central loop. RNA 20: 1129-1141

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

Jones, J.D., Shlumukov, L., Carland, F., English, J., Scofield, S.R., Bishop, G.J., Harrison, K. 1992. Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. Transgenic Research 1: 285-297

Jung, C. and Müller, A.E. 2009. Flowering time control and applications in plant breeding. Trends in Plant Science 14: 563-573

Jung, J.H., Park, J.H., Lee, S., To, T.K., Kim, J.M., Seki, M., Park, C.M. 2013. The cold signaling attenuator HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE1 activates *FLOWERING LOCUS C* transcription via chromatin remodeling under short-term cold stress in *Arabidopsis*. The Plant Cell 25: 4378-4390

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316: 1484-1488

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H. 2005. Antisense transcription in the mammalian transcriptome. Science 309: 1564-1566

Katinakis, P.K., Slater, A., Burdon, R.H. 1980. Non-polyadenylated mRNAs from eukaryotes. FEBS Letters 116: 1-7

Kemi, U., Niittyvuopio, A., Toivainen, T., Pasanen, A., Quilot-Turion, B., Holm, K., Lagercrantz, U., Savolainen, O., Kuittinen, H. 2013. Role of vernalization and of duplicated *FLOWERING LOCUS C* in the perennial *Arabidopsis lyrata*. New Phytologist 197: 323-335

Khavari, D.A., Sen, G.L., Rinn, G.L. 2010. DNA methylation and epigenetic control of cellular differentiation. Cell Cycle 9: 3880-3883

Kim, D.H. and Sung, S. 2017. Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. Developmental Cell 40: 302-312

Kim, S., Choi, K., Park, C., Hwang, H-J., Lee, I. 2006. *SUPPRESSOR OF FRIGIDA4*, encoding a C2H2-type zinc finger protein, represses flowering by transcriptional activation of *Arabidopsis FLOWERING LOCUS C*. The Plant Cell 18: 2985-2998

Kim, S.H., Sussman, J.L., Suddath, F.L., Quigley, G.J., McPherson, A., Wang, A.H.J., Seeman, N.C., Rich, A. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185: 435-440

Kim, S.Y. and Michaels, S.D. 2006. *SUPPRESSOR OF FRI4* encodes a nuclear-localized protein that is required for delayed flowering in winter-annual *Arabidopsis*. Development 133: 4699-4707

Kim, S-Y., Park, B-S., Kwon, S-J., Kim, J., Lim, M-H., Park, Y-D., Kim, D.Y., Suh, S-C., Jin, Y-M., Ahn, J.H., Lee, Y-H. 2007. Delayed flowering time in *Arabidopsis* and *Brassica rapa* by the overexpression of *FLOWERING LOCUS C* (*FLC*) homologs isolated from Chinese Cabbage (*Brassica rapa* L. ssp. *pekinensis*). Plant Cell Reports 26: 327-336

Koch, M.A. and Kiefer, M. 2005. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species—*Capsella rubella*, *Arabidopsis lyrata* subsp. petraea, and *A. thaliana*. American Journal of Botany 92: 761-767

Kole, C., Quijada, P., Michaels, S.D., Amasino, R.M., Osborn, T.C. 2001. Evidence for homology of flowering-time genes *VFR2* from *Brassica rapa* and *FLC* from *Arabidopsis thaliana*. Theoretical and Applied Genetics 102: 425-430

Koornneef, M., Alonso-Blanco, C., Blankestijn-de Vries, H., Hahart, C.J., Peeters, A.J.M. 1998. Genetic Interactions among late-flowering mutants of *Arabidopsis*. Genetics 148: 885-892

Koornneef, M., Blankestijn-de Vries, H., Hanhart, C., Soppe, W., Peeters, T. 1994. The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg *erecta* wild-type. The Plant Journal 6: 911-919

Koornneef, M., Hanhart, C.J., Van der Veen, J.H. 1991. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana.* Molecular Genetics and Genomics 229: 57-66

Kornblihtt, A.R. 2007. Coupling transcription and alternative splicing. Advances in Experimental Medicine and Biology 623: 175-189

Kortmann, J. and Narberhaus, F. 2012. Bacterial RNA thermometers: molecular zippers and switches. Nature Reviews Microbiology 10: 255-265

Kuittinen, H., Niittyvuopio, A., Rinne, P., Savolainen, O. 2007. Natural variation in *Arabidopsis lyrata* vernalization requirement conferred by a *FRIGIDA* indel polymorphism. Molecular Biology and Evolution 25: 319-329

Kumar, G., Arya, P., Gupta, K., Randhawa, V., Acharya, V., Singh, A.K. 2016. Comparative phylogenetic analysis and transcriptional profiling of MADS-box gene family identified *DAM* and *FLC*-like genes in apple (*Malusx domestica*). Scientific Reports 6: 20695

Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T., Marques, A.C. 2012. Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genetics 8: e1002841

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., Reinberg, D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes & Development 16: 2893-2905

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. The Plant Journal 9: 13-20

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294: 853-858

Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W., Jonak, C., Hauser, M.T. and Luschnig, C. 2010. Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in *Arabidopsis*. Molecular Plant 3: 594-602

Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294: 858-862

Le Corre, V., Roux, F., Reboud, X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Molecular Biology and Evolution 19: 1261-1271

Lee, H., Suh, S-S., Park, F., Cho, E., Ahn, J.H., Kim, S-G., Lee, J.S., Kwon, Y.M., Lee, I. 2000. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. Genes & Development 14: 2366-2376

Lee, I., Aukerman, M.J., Gore, S.L., Lohman, K.N., Michaels, S.D., Weaver, L.M., John, C.M., Feldmann, K.A., Amasino, R.M. 1994**A**. Isolation of *LUMNIDEPENDENS*: a gene involved in the control of flowering time in *Arabidopsis*. The Plant Cell 6: 75-83

Lee, I., Michaels, S.D., Masshardt, A.S., Amasino, R.M. 1994**B**. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of *Arabidopsis*. The Plant Journal 6: 903-909

Lee, J. and Amasino, R.M. 2013. Two *FLX* family members are non-redundantly required to establish the vernalization requirement in *Arabidopsis*. Nature Communications 4: 2186

Lee, J., Oh, M., Park, H., Lee, I. 2008. SOC1 translocated to the nucleus by interaction with AGL24 directly regulates *LEAFY*. The Plant Journal 55: 832-843

Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., Ahn, J.H. 2007. Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Genes & Development 21: 397-402

Lee, R.C. and Ambros, V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294: 862-864

Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R. Dean, C. 2002. Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. Science 297: 243-246

Lewis, C.J., Pan, T., Kalsotra, A. 2017. RNA modifications and structures cooperate to guide RNA-protein interactions. Nature Reviews Molecular Cell Biology, 18: 202-210

Life Technologies, 2012. Real-time PCR handbook. Life Technologies Corporation

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., Yu, H. 2008. A repressor complex governs the integration of flowering signals in *Arabidopsis*. Developmental Cell 15: 110-120

Li, P., Filiault, D., Box, M.S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A.M., Schmitt, J., McMullan, M., Bergelson, J., Nordborg, M., Dean, C. 2014. Multiple *FLC* haplotypes defined by independent *cis*-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. Genes & Development 28: 1635-1640

Li, P., Tao, Z., Dean, C. 2015. Phenotypic evolution through variation in splicing of the noncoding RNA *COOLAIR*. Genes & Development 29: 696-701

Li, X., Zhang, S., Bai, J., He, Y. 2016. Tuning growth cycles of *Brassica* crops via natural antisense transcripts of *BrFLC*. Plant Biotechnology Journal 14: 905-914

Lim, M-H., Kim, J., Kim, Y-S., Chung, K-S., Seo, Y-H., Lee, I., Kim, J., Hong, C.B., Kim, H-J., Park, C-M. 2004. A new *Arabidopsis* gene, *FLK*, encodes an RNA binding protein with K homology motifs and regulates flowering time via *FLOWERING LOCUS C*. The Plant Cell 16: 731-740

Lin, S-I., Wang, J-G., Poon, S-Y., Su, C-l., Wang, S-S., Chiou, T-J. 2005. Differential regulation of *FLOWERING LOCUS C* expression by vernalization in cabbage and *Arabidopsis*. Plant Physiology 137: 1037-1048

Lipps, H.J. and Rhodes, D. 2009. G-quadruplex structures: *in vivo* evidence and function. Trends in Cell Biology 19: 414-422

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., Dean, C. 2010. Targeted 3′ processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. Science 327: 94-97

Liu, F., Quesada, V., Crevillén, P., Bäurle, I., Swiezewski, S. and Dean, C., 2007. The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*. Molecular Cell 28: 398-407

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., Pan, T. 2015. N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature 518: 560-564

Livak, K.J. and Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25: 402-408

Lloyd, A. and Bomblies, K. 2016. Meiosis in autopolyploid and allopolyploid *Arabidopsis*. Current Opinion in Plant Biology 30: 116-122

Lokody, I. 2014. RNA: RiboSNitches reveal heredity in RNA secondary structure. Nature Reviews Genetics 15: 219-219

Long, Y., Shi, J., Qiu, D., Li, R., Zhang, C., Wang, J., Hou, J., Zhao, J., Shi, L., Park, B-S., Choi, S.R., Lim, Y.P., Meng, J. 2007. Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis*. Genetics 177: 2433-2444

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

Lou, P., Zhao, J., Kim, J.S., Shen, S., Carpio, D.P.D., Song, X., Jin, M., Vreugdenhil, D., Wang, X., Koornneef, M., Bonnema, G. 2007. Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. Journal of Experimental Botany 58: 4005-4016

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. Cell 165: 1267-1279

Lukens, L. N., Quijada, P. A., Udall, J., Pires, J. C., Schranz, M. E., Osborn, T. C. 2004. Genome redundancy and plasticity within ancient and recent *Brassica* crop species. Biological Journal of the Linnean Society 82: 665–674

Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1155

Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., Schubert, I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. PNAS 103: 5224-5229

Lysak, M.A., Koch, M.A., Pecinka, A., Schubert, I. 2005. Chromosome triplication found across the tribe Brassiceae. Genome Research 15: 516-525

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y., Xie, Y. 2015. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Molecular Plant 8: 1274-1284

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J. and Seymour, G.B. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948-952

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L, Murphy, G., Sherson, S., Cobbett, C., Dean, C. 1997. *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. Cell 89: 737-745

Macknight, R., Duroux, M., Laurie, R., Dikjwel, P., Simpson, G., Dean, C. 2002. Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter *FCA*. The Plant Cell 14: 877-888

Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q., Dean, C. 2014. Functional consequences of splicing of the antisense transcript *COOLAIR* on *FLC* transcription. Molecular Cell 54: 156-165

Marques, A.C. and Ponting, C.P. 2009. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biology 10: R124

Martens, J.A., Laprade, L., Winston, F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae SER3* gene. Nature 429: 571-574

Martin, J.S., Halvorsen, M., Davis-Neulander, L., Ritz, J., Gopinath, C., Beauregard, A., Laederach, A. 2012. Structural effects of linkage disequilibrium on the transcriptome. RNA 18: 77-87

Mattick, J.S. 2011. The central role of RNA in human development and cognition. FEBS Letters 585: 1600-1616

Matzke, A.J. and Matzke, M.A. 1998. Position effects and epigenetic silencing of plant transgenes. Current Opinion in Plant Biology 1: 142-148

Meister, A. 2005. Calculation of binding length of base-specific DNA dyes by comparison of sequence and flow cytometric data. Application to *Oryza sativa* and *Arabidopsis thaliana*. Journal of Theoretical Biology 232: 93-97

Melé, M. and Rinn, J.L. 2016. "Cat's Cradling" the 3D genome by the act of lncRNA transcription. Molecular Cell 62: 657-664

Méndez-Vigo, B., Savic, M., Ausín, I., Ramiro, M., Martín, B., Picó, F.X., Alonso-Blanco, C. 2016. Environmental and genetic interactions reveal *FLOWERING LOCUS C* as a modulator of the natural variation for the plasticity of flowering in *Arabidopsis*. Plant, Cell & Environment 39: 282-294

Merino, E.J., Wilkinson, K.A., Coughlan, J.L, Weeks, K.M. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). Journal of the American Chemical Society 127: 4223-31

Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E., Vilardell, J. 2011. Deciphering 3'ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. Molecular Cell 43: 1033-1039

Michaels, S., Bezerra, I., Amasino, R. 2004. *FRIGIDA*-related genes are required for the winter-annual habit in *Arabidopsis*. PNAS 101: 3281-3285

Michaels, S.D. and Amasino, R.M., 1999. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell 11: 949-956

Michaels, S.D., Amasino, R.M. 2001. Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. The Plant Cell 13: 935-941

Michaels, S.D., He, Y., Scortecci, K.C., Amasino, R.M. 2003. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. PNAS 100: 10102-10107

Montange, R.K. and Batey, R.T. 2008. Riboswitches: emerging themes in RNA structure and function. Annual Review of Biophysics 37: 117-133

Montemayor, E.J., Curran, E.C., Liao, H.H., Andrews, K.L., Treba, C.N., Butcher, S.E., Brow, D.A. 2014. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. Nature Structural & Molecular Biology 21: 544-551

Moon, J., Suh, S-S., Horim, L., Choi, K-R., Hong, C.B., Paek, N-C., Kim, S-G., Lee, I. 2003. The *SOC1* MADS-box gene integrates vernalization and gibberellin signals for flowering in *Arabidopsis*. The Plant Journal 35: 613-623

Moore, M.J. and Proudfoot, N.J. 2009. Pre-mRNA processing reaches back totranscription and ahead to translation. Cell 136: 688-700

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., Dreyfuss, G. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes & Development 16: 720-728

Mukundan, V.T., Phan, A.T. 2013. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. Journal of the American Chemical Society 135: 5017-5028

Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Crollius, H.R., Salse, J. 2015. Understanding Brassicaceae evolution through ancestral genome reconstruction. Genome Biology 16: 262

Murray, M.G. and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8: 4321-4326

Mutasa-Göttgens, E. and Hedden, P. 2009. Gibberellin as a factor in floral regulatory networks. Journal of Experimental Botany 60: 1979-1989

Mylne, J.S., Barrett, L., Tessadori, F., Mesnage, S., Johnson, L., Bernatavichute, Y.V., Jacobsen, S.E., Fransz, P., Dean, C. 2006. LHP1, the *Arabidopsis* homologue of HETEROCHROMATIN PROTEIN 1, is required for epigenetic silencing of *FLC*. PNAS 103: 5012-5017

Nagaharu, U. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. The Journal of Japanese Botany 7: 389-452

Nakamura, M. and Hennig, L. 2017. Inheritance of vernalization memory at *FLOWERING LOCUS C* during plant regeneration. Journal of Experimental Botany 68: 2813-2819

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., Zakian, S.M. 2001. Characterization of the genomic *Xist* locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Research 11: 833-849

Nguyen, L.A., Wang, J., Steitz, T.A. 2017. Crystal structure of Pistol, a class of self-cleaving ribozyme. PNAS 114: 1021-1026

Nilsen, T.W. and Graveley, B.R. 2010. Expansion of the eukaryotic proteome by alternative splicing. Nature 463: 457-463

Nilsson, O., Lee, I., Blázquez, M.A., Weigel, D. 1998. Flowering-time genes modulate the response to *LEAFY* activity. Genetics 150: 403-410

Nitsche, A., Rose, D., Fasold, M., Reiche, K., Stadler, P.F. 2015. Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved. RNA 21: 801-812

Noller, H.F. and Woese, C.R. 1981. Secondary structure of 16S ribosomal RNA. Science 212: 403-411

Novikova, I.V., Dharap, A., Hennelly, S.P., Sanbonmatsu, K.Y. 2013. 3S: shotgun secondary structure determination of long non-coding RNAs. Methods 63: 170-177

Novikova, I.V., Hennelly, S.P., Sanbonmatsu, K.Y. 2012. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Research 40: 5034-5051

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*. Theoretical and Applied Genetics 114: 595-608

Osborn, T.C., Kole, C., Parkin, I.A.P., Sharpe, A.G., Kuiper, M., Lydiate, D., Trick, M. 1997. Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics 146: 1123-1129

Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H-S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., Martienssen, R.A. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19: 141-147

Palumbo, S.L., Ebbinghaus, S.W., Hurley, L.H. 2010. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. Journal of the American Chemical Society 131: 10878-10891

Pang, K.C., Frith, M.C., Mattick, J.S. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends in Genetics 22: 1-5

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., Poethig, R.S. 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. PNAS 102: 3691-3696

Parkin, I.A.P., Gulden, S.M., Sharpe, A.G., Lukens, L., Trick, M., Osborn, T.C., Lydiate, D.J. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171: 765-781

Pelechano, V. and Steinmetz, L.M. 2013. Gene regulation by antisense transcription. Nature Reviews Genetics 14: 880-893

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N. 1996. Requirement for *Xist* in X chromosome inactivation. Nature 379: 131-137

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., Mazo, A. 2006. Transcriptional elongation of non-coding *bxd* RNAs promoted by the Trithorax TAC1 complex represses *Ubx* by a transcriptional interference mechanism. Cell 127: 1209-1221

Pires, J.C., Zhao, J., Schranz, M.E., Leon, E.J., Quijada, P.A., Lukens, L.N., Osborn, T.C. 2004. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). Biological Journal of the Linnean Society 82: 675-688

Ponjavic, J., Ponting, C.P., Lunter, G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Research 17: 556-565

Putterill, J., Robson, F., Lee, K., Simon, R., Coupland, G. 1995. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80: 847-857

Queseda, V., Macknight, R., Dean, C., Simpson, G.G. 2003. Autoregulation of *FCA* pre-mRNA processing controls *Arabidopsis* flowering time. The EMBO Journal 22: 3142-3152

Qüesta, J.I., Song, J., Geraldo, N., An, H., Dean, C. 2016. *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at *FLC* during vernalization. Science 353: 485-488

Quinn, J.J., Zhang, Q.C., Georgiev, P., Ilik, I.A., Akhtar, A., Chang, H.Y. 2016. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. Genes & Development 30: 191-207

Rafferty, N.E. and Nabity, P.D. 2017. A global test for phylogenetic signal in shifts in flowering time under climate change. Journal of Ecology 105: 627-633

Rana, D., van den Boogaart, T., O'Neill1, C.M., Hynes, L., Bent, E., Macpherson, L., Park, J.Y., Lim, Y.P., Bancroft, I. 2004. Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. The Plant Journal 40: 725-733

Ratcliffe, O.J., Kumimoto, R.W., Wong, B.J., Riechmann, J.L. 2003. Analysis of the *Arabidopsis MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. The Plant Cell 15: 159-1169.

Razi, H., Howell, E.C., Newbury, H.J., Kearsey, M.J. 2008. Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*? Theoretical and Applied Genetics 116: 179-192

Reed, R. and Maniatis, T. 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell 46: 681-690

Reeves, P.A., He, Y., Schmitz, R.J., Amasino, R.M., Panella, L.W., Richards, C.M. 2007. Evolutionary conservation of the *FLOWERING LOCUS C*-mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). Genetics 176: 295-307

Ren, A., Vušurović, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., Micura, R., Patel, D.J. 2016. Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. Nature Chemical Biology: 702-708

Ridge, S., Brown, P.H., Hecht, V., Driessen, R.G. and Weller, J.L. 2015. The role of *BoFLC*2 in cauliflower (*Brassica oleracea* var. botrytis L.) reproductive development. Journal of Experimental Botany 66: 125-135

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., Chang, H.Y. 2007. Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. Cell 129: 1311-1323

Ritz, J., Martin, J.S., Laederach, A. 2012. Evaluating our ability to predict the structural disruption of RNA by SNPs. BMC Genomics 13: 56

Rivas, E., Clements, J., Eddy, S.R. 2016. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nature Methods: 45-48

Rosa, S., Duncan, S., Dean, C. 2016. Mutually exclusive sense–antisense transcription at *FLC* facilitates environmentally induced gene repression. Nature Communications 7: 13031

Roth, A. and Breaker, R.R. 2009. The structural and functional diversity of metabolite-binding riboswitches. Annual Review of Biochemistry 78: 305-334

Ruelens, P., De Maagd, R.A., Proost, S., Theißen, G., Geuten, K., Kaufmann, K. 2013. *FLOWERING LOCUS C* in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nature Communications 4: 2280

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., Coupland, G. 2000. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science 2888: 1613-1616

Sánchez-Bermejo, E., Méndez-Vigo, B., Pico, F.X., Martínez-Zapater, J.M., Alonso-Blanco, C. 2012. Novel natural alleles at *FLC* and *LVR* loci account for enhanced vernalization responses in *Arabidopsis thaliana*. Plant, Cell & Environment 35: 1672-1684

Sanda, S., John, M., Amasino, R. 1997. Analysis of flowering time in ecotypes of *Arabidopsis thaliana*. Journal of Heredity 88: 69-72

Sanda, S.L. and Amasino, R.M. 1996. Interaction of *FLC* and late-flowering mutations in *Arabidopsis thaliana*. Molecular and General Genetics MGG 251: 69-74

Satake, A., Kawagoe, T., Saburi, Y., Chiba, Y., Sakurai, G., Kudoh, H. 2013. Forecasting flowering phenology under climate warming by modelling the regulatory dynamics of flowering-time genes. Nature Communications 4: 2303

Sawa, M., Nusinow, D.A., Kay, S.A., Imaizumi, T. 2007. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. Science 318: 261-265

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

Schiessl, S., Samans, B., Hüttel, B., Reinhard, R., Snowdon, R.J. 2014. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. Frontiers in Plant Science 5: 404

Schmitz, R.J., Hong, L., Fitzpatrick, K.E., Amasino, R.M. 2007. DICER-LIKE 1 and DICER-LIKE 3 redundantly act to promote flowering via repression of *FLOWERING LOCUS C* in *Arabidopsis thaliana*. Genetics 176: 1359-1362

Schmitz, R.J., Hong, L., Michaels, S., Amasino, R.M. 2005. *FRIGIDA-ESSENTIAL 1* interacts genetically with *FRIGIDA* and *FRIGIDA-LIKE 1* to promote the winter-annual habit of *Arabidopsis thaliana*. Development 132: 5471-5478

Schnable, J.C., Springer, N.M., Freeling, M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. PNAS 108: 4069-4074

Schomburg, F.M., Patton, D.A., Meinke, D.W., Amasino, R.M. 2001. *FPA*, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs. The Plant Cell 13: 1427-1436

Schorderet, P. and Duboule, D. 2011. Structural and functional differences in the long non-coding RNA *Hotair* in mouse and human. PLoS Genetics 7: e1002071

Schranz, M.E., Lysak, M.A., Mitchell-Olds, T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends in Plant Science 11: 535-542

Schranz, M.E., Quijada, P., Sung, S-B., Lukens, L., Amasino, R., Osborn, T.C. 2002. Characterization and effects of the replicated flowering time gene *FLC* in *Brassica rapa*. Genetics 162: 1457-1468

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., Goodrich, J. 2006. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. The EMBO Journal 25: 4638-4649

Shearwin, K.E., Callen, B.P., Egan, J.B. 2005. Transcriptional interference–a crash course. TRENDS in Genetics 21: 339-345

Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J., Dennis, E.S. 1999. The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulation by vernalization and methylation. The Plant Cell 11: 445-458

Sheldon, C.C., Conn, A.B., Dennis, E.S., Peacock, W.J. 2002. Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. The Plant Cell 14: 2527-2537

Sheldon, C.C., Rouse, D.T., Finnegan, E.J., Peacock, W.J., Dennis, E.S. 2000. The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). PNAS 97: 3753-3758

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., Shi, Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119: 941-953

Shindo, C., Aranzana, M.J., 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 Physiology 138: 1163-1173

Shindo, C., Lister, C., Crevillen, P., Nordborg, M., Dean, C. 2006. Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response. Genes & Development 20: 3079-3083

Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C.Y., Su, Y., Xia, X., Yin, W. 2014. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. Journal of Experimental Botany 65: 4975-4983

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M., Van de Peer, Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. PNAS 99: 13627-13632

Simpson, G.G., Dijkwel, P.P., Queseda, V., Henderson, I., Dean, C. 2003. FY is an RNA 3'end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. Cell 113: 777-787

Simpson, G.G., Laurie, R.E., Dijkwel, P.P., Quesada, V., Stockwell, P.A., Dean, C., Macknight, R.C. 2010. Noncanonical translation initiation of the *Arabidopsis* flowering time and alternative polyadenylation regulator FCA. The Plant Cell 22: 3764-3777

Skourti-Stathaki, K., Kamieniarz-Gdula, K., Proudfoot, N.J. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature 516: 436-439

Smith, C.W., Chu, T.T., Nadal-Ginard, B. 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Molecular and Cellular Biology 13: 4939-4952

Smith, M.A., Gesell, T., Stadler, P.F., Mattick, J.S. 2013. Widespread purifying selection on RNA structure in mammals. Nucleic Acids Research 41: 8220-8236

Snowdon, R.J. and Friedt, W. 2004. Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. Plant Breeding 123: 1-8

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., Pyle, A.M. 2015. *HOTAIR* forms an intricate and modular secondary structure. Molecular Cell 58: 353-361

Song, J., Irwin, J., Dean, C. 2013**B**. Remembering the prolonged cold of winter. Current Biology 23: R807-R811

Song, K., Lu, P.I.N.G., Tang, K., Osborn, T.C. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. PNAS 92: 7719-7723

Song, Y.H., Ito, S., Imaizumi, T. 2013**A**. Flowering time regulation: photoperiod-and temperature-sensing in leaves. Trends in Plant Science 18: 575-583

Sonmez, C., Bäurle, I., Magusin, A., Dreos, R., Laubinger, S., Weigel, D., Dean, C. 2011. RNA 3' processing functions of *Arabidopsis* FCA and FPA limit intergenic transcription. PNAS 108: 8508-8513

Stavropoulos, N., Lu, N., Lee, J.T. 2001. A functional role for *Tsix* transcription in blocking *Xist* RNA accumulation but not in X-chromosome choice. PNAS 98: 10232-10237

Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., Purugganan, M.D., Schmitt, J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. PNAS 101: 4712-4717

Strange, A., Li, P., Lister, C., Anderson, J., Warthmann, N., Shindo, C., Irwin, J., Nordborg, M., Dean, C. 2011. Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions. PLoS ONE 6: e19949

Struhl, K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nature Structural & Molecular Biology 14: 103-105

Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H. 2001. *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410: 1116-1120

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., Dean, C. 2013. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. Science 340: 619-621

Sung, S. and Amasino, R.M. 2004. Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. Nature 427: 159-164

Sung, S., He, Y., Eshoo, T.W., Tamada, Y., Johnson, L., Nakahigashi, K., Goto, K., Jacobsen, S.E., Amasino, R.M. 2006. Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires LIKE HETEROCHROMATIN PROTEIN 1. Nature Genetics 38: 706-710

Swiezewski, S., Crevillén, P., Liu, F., Ecker, J.R., Jerzmanowski, A., Dean, C. 2007. Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, *FLC*. PNAS 104: 3633-3638

Swiezewski, S., Liu, F., Magusin, A., Dean, C.,2009. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. Nature 462: 799-802

Tadege, M., Sheldon, C.C., Helliwell, C.A., Stoutjesdijk, P., Dennis, E.S., Peacock, W.J. 2001. Control of flowering time by *FLC* orthologues in *Brassica napus*. The Plant Journal 28: 545-553

Tang, H., Woodhouse, M.R., Cheng, F., Schnable, J.C., Pedersen, B.S., Conant, G., Wang, X., Freeling, M., Pires, J.C. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. Genetics 190: 1563-1574

Taylor, J.L., Massiah, A., Kennedy, S., Hong, Y., Jackson, S.D. 2017. *FLC* expression is down-regulated by cold treatment in *Diplotaxis tenuifolia* (wild rocket), but flowering time is unaffected. Journal of Plant Physiology 214: 7-15

Tinoco, I. and Bustamante, C. 1999. How RNA folds. Journal of Molecular Biology 293: 271-281

Tiwari, S.B., Shen, Y., Chang, H.C., Hou, Y., Harris, A., Ma, S.F., McPartland, M., Hymus, G.J., Adam, L., Marion, C., Belachew, A. 2010. The flowering time regulator CONSTANS is recruited to the *FLOWERING LOCUS T* promoter via a unique *cis*-element. New Phytologist 187: 57-66

Town, C.D., Cheung, F., Maiti, R., Crabtree, J., Haas, B.J., Wortman, J.R., Hine, E.E., Althoff, R., Arbogast, T.S., Tallon, L.J., Vigouroux, M. 2006. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. The Plant Cell 18: 1348-1359

Traub, W. and Sussman, J.L. 1982. Adenine-guanine base pairing in ribosomal RNA. Nucleic Acids Research 10: 2701-2708

Treisman, R., Proudfoot, N.J., Shander, M., Maniatis, T. 1982. A single-base change at a splice site in a β0-thalassemic gene causes abnormal RNA splicing. Cell 29: 903-911

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., Chang, H.Y. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. Science 329: 689-693

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., Colot, V. 2007. *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genetics 3: e86.

Udall, J.A., Quijada, P.A., Osborn, T.C. 2005. Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. Genetics 169: 967-979

Valentim, F.L., van Mourik, S., Posé, D., Kim, M.C., Schmid, M., van Ham, R.C., Busscher, M., Sanchez-Perez, G.F., Molenaar, J., Angenent, G.C., Immink, R.G. 2015. A quantitative and dynamic model of the *Arabidopsis* flowering time gene regulatory network. PloS ONE 10: e0116973

Vision, T.J., Brown, D.J., Tanksley, S.D. 2000. The origins of genomic duplications in *Arabidopsis*. Science 290: 2114-2117

Walsh, J.B. 1995. How often do duplicated genes evolve new functions? Genetics 139: 421-428

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., Chang, H.Y. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. Nature 505: 706-709

Wang, H., Chung, P.J., Liu, J., Jang, I.C., Kean, M.J., Xu, J., Chua, N.H. 2014**B**. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. Genome Research 24: 444-453

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., Wong, G.K.S. 2004. Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. Nature 431: doi:10.1038/nature03016

Wang, J.W., Czech, B., Weigel, D. 2009**A**. miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. Cell 138: 738-749

Wang, N., Qian, W., Suppanz, I., Wei, L., Mao, B., Long, Y., Meng, J., Müller, A.E., Jung, C. 2011. Flowering time variation in oilseed rape (*Brassica napus* L.) is associated with allelic variation in the *FRIGIDA* homologue *BnaA.FRI.a*. Journal of Experimental Botany 62: 5641-5658

Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., Alonso-Blanco, C., Coupland, G., Albani, M.C. 2009**B**. *PEP1* regulates perennial flowering in *Arabis alpina*. Nature 459: 423-427

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nature Genetics 43: 1035-1039

Wang, Z., Zhang, X.J., Ji, Y.X., Zhang, P., Deng, K.Q., Gong, J., Ren, S., Wang, X., Chen, I., Wang, H., Gao, C. 2016. The long noncoding RNA *Chaer* defines an epigenetic checkpoint in cardiac hypertrophy. Nature Medicine 22: 1131-1139

Wang, Z.W., Wu, Z., Raitskin, O., Sun, Q., Dean, C. 2014**A**. Antisense-mediated *FLC* transcriptional repression requires the P-TEFb transcription elongation factor. PNAS 111: 7468-7473

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., Meyerowitz, E.M. 1992. *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell 69: 843-859

Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., Neph, S. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Research 35: 4809-4819

Werner, J.D., Borevitz, J.O., Uhlenhaut, N.H., Ecker, J.R., Chory, J., Weigel, D. 2005**A**. *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genetics 170: 1197-1207

Werner, J.D., Borevitz, J.O., Warthmann, N., Trainer, G.T., Ecker, J.R., Chory, J., Weigel, D. 2005**B**. Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. PNAS 102: 2460-2465

Wierzbicki, A.T., Ream, T.S., Haag, J.R., Pikaard, C.S. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature Genetics 41: 630-634

Wilson, R.N., Heckman, J.W., Somerville, C.R. 1992. Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. Plant Physiology 100: 403-408

Wollenberg, A.C. and Amasino, R.M. 2012. Natural variation in the temperature range permissive for vernalization in accessions of *Arabidopsis thaliana*. Plant, Cell & Environment 35: 2181-2191

Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., Helliwell, C.A. 2006. The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. PNAS 103: 14631-14636

Woodhouse, M.R., Cheng, F., Pires, J.C., Lisch, D., Freeling, M., Wang, X. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. PNAS 111: 5283-5288

Wu, J., Okada, T., Fukushima, T., Tsudzuki, T., Sugiura, M., Yukawa, Y. 2012**B**. A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in *Arabidopsis*. RNA Biology 9: 302-313

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

Xiao, D., Zhao, J.J., Hou, X.L., Basnet, R.K., Carpio, D.P., Zhang, N.W., Bucher, J., Lin, K., Cheng, F., Wang, X.W., Bonnema, G. 2013. The *Brassica rapa FLC* homologue *FLC*2 is a key regulator of flowering time, identified through transcriptional co-expression networks. Journal of Experimental Botany 64: 4503-4516

Xin, M., Wang, Y., Yao, Y., Song, N., Hu, Z., Qin, D., Xie, C., Peng, H., Ni, Z., Sun, Q. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biology 11: 61

Xue, Z., Hennelly, S., Doyle, B., Gulati, A.A., Novikova, I.V., Sanbonmatsu, K.Y., Boyer, L.A. 2016. A G-rich motif in the lncRNA *Braveheart* interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. Molecular Cell 64: 37-50

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302: 842-846

Yamaguchi, A., Wu, M.F., Yang, L., Wu, G., Poethig, R.S., Wagner, D. 2009. The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of *LEAFY*, *FRUITFULL*, and *APETALA1*. Developmental Cell 17: 268-278

Yang, H., Howard, M., Dean, C. 2014. Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at *Arabidopsis FLC*. Current Biology 24: 1793-1797

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K.A., Grinstein, J.D., Dorrestein, P.C., Rosenfeld, M.G. 2011. ncRNA-and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell 147: 773-788

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., Zhou, M.M. 2010. Molecular interplay of the noncoding RNA *ANRIL* and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of *INK4a*. Molecular Cell 38: 662-674

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S. 2003. Widespread occurrence of antisense transcription in the human genome. Nature Biotechnology 21: 379

Yi, G., Park, H., Kim, J.S., Chae, W.B., Park, S., Huh, J.H. 2014. Identification of three *FLOWERING LOCUS C* genes responsible for vernalization response in radish (*Raphanus sativus* L.). Horticulture, Environment, and Biotechnology 55: 548-556

Yuan, Y-X., Wu, J., Sun, R-F., Zhang, X-W., Xu, D-H., Bonnema, G., Wang, X-W. 2009. A naturally occurring splicing site mutation in the *Brassica rapa FLC1* gene is associate with variation in flowering time. Journal of Experimental Botany 60: 1299-1308

Zhang, W., Han, Z., Guo, Q., Liu, Y., Zheng, Y., Wu, F., Jin, W. 2014. Identification of maize long non-coding RNAs responsive to drought stress. PLoS ONE 9: e98958

Zhang, Z., Lu, J., Cong, R., Ren, T., Li, X. 2017. Evaluating agroclimatic constraints and yield gaps for winter oilseed rape (*Brassica napus* L.)–A case study. Scientific Reports 7: 7852

Zhao, J., Kulkami, V., Liu, N., Carpio, D.P.D., Bucher, J., Bonnema, G. 2010. *BrFLC2* (*FLOWERING LOCUS C*) as a candidate gene for a vernalization response QTL in *Brassica rapa*. Journal of Experimental Botany 61: 1817-1825

Zhu, Q.H., Stephen, S., Taylor, J., Helliwell, C.A., Wang, M.B. 2014. Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. New Phytologist 201: 574-584

Zou, X., Suppanz, I., Raman, H., Hou, J., Wang, J., Long, Y., Jung, C., Meng, J. 2012. Comparative analysis of *FLC* homologues in Brassicaceae provides insight into their role in the evolution of oilseed rape. PLoS ONE 7: e45751

Zubko, E. and Meyer, P. 2007. A natural antisense transcript of the *Petunia hybrida Sho* gene suggests a role for an antisense mechanism in cytokinin regulation. The Plant Journal 52: 1131-1139

8 APPENDICES

8.1 SUPPLEMENTARY FIGURES

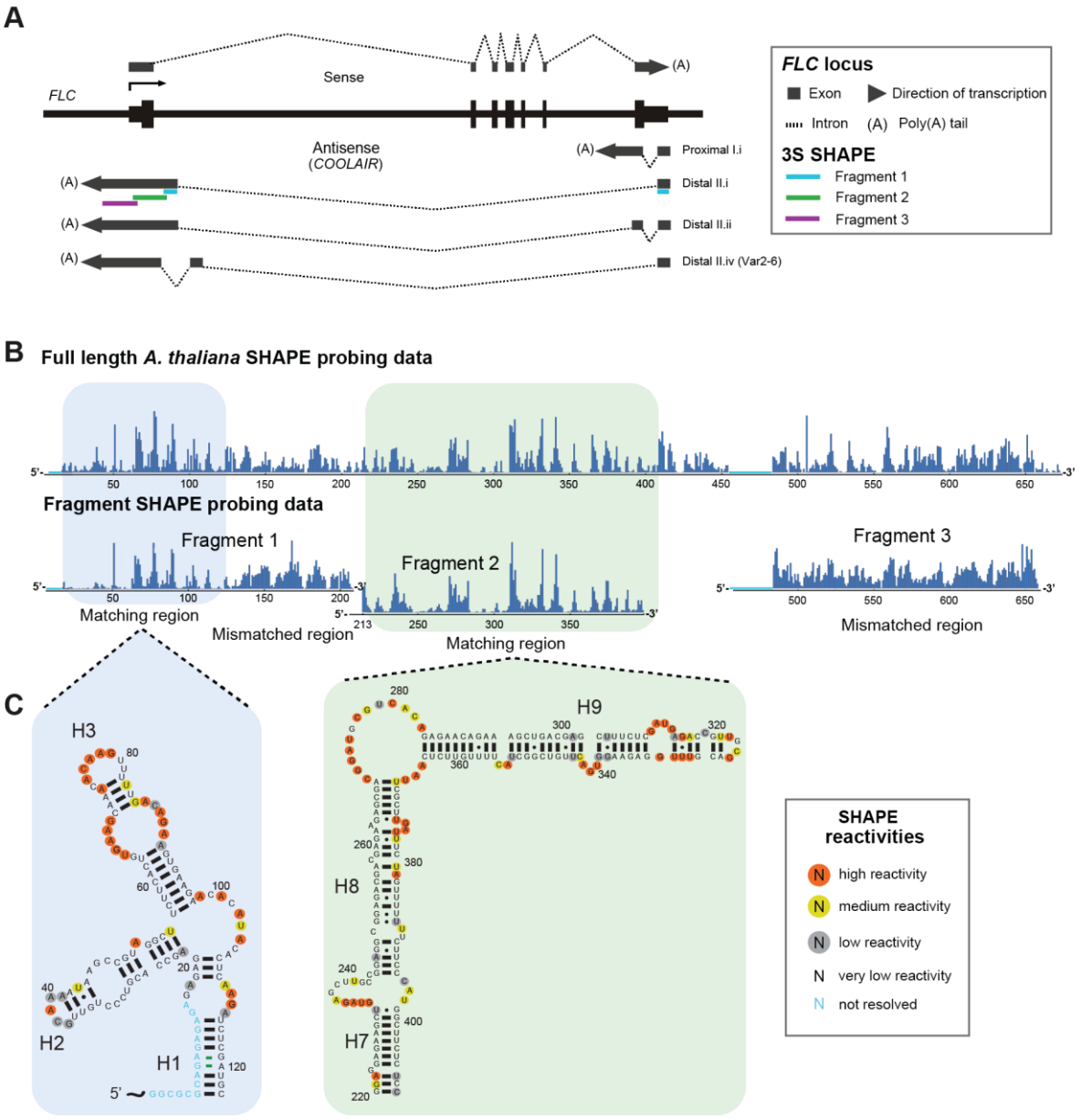


Figure S1: 3S determination of the *A. thaliana* distal *COOLAIR* secondary structure. (A) Schematic representation of *FLC* and *COOLAIR* transcripts at the *FLC* locus, with 3S fragment positions mapped. (B) SHAPE reactivities for the full-length *A. thaliana* distal *COOLAIR* (Class II.i) transcript are compared with shorter fragments 1-3 for 3S determination. (C) Modular secondary structure corresponding to reactivity data of the boxed regions in (B). Figure reproduced from Hawkes *et al.* (2016).

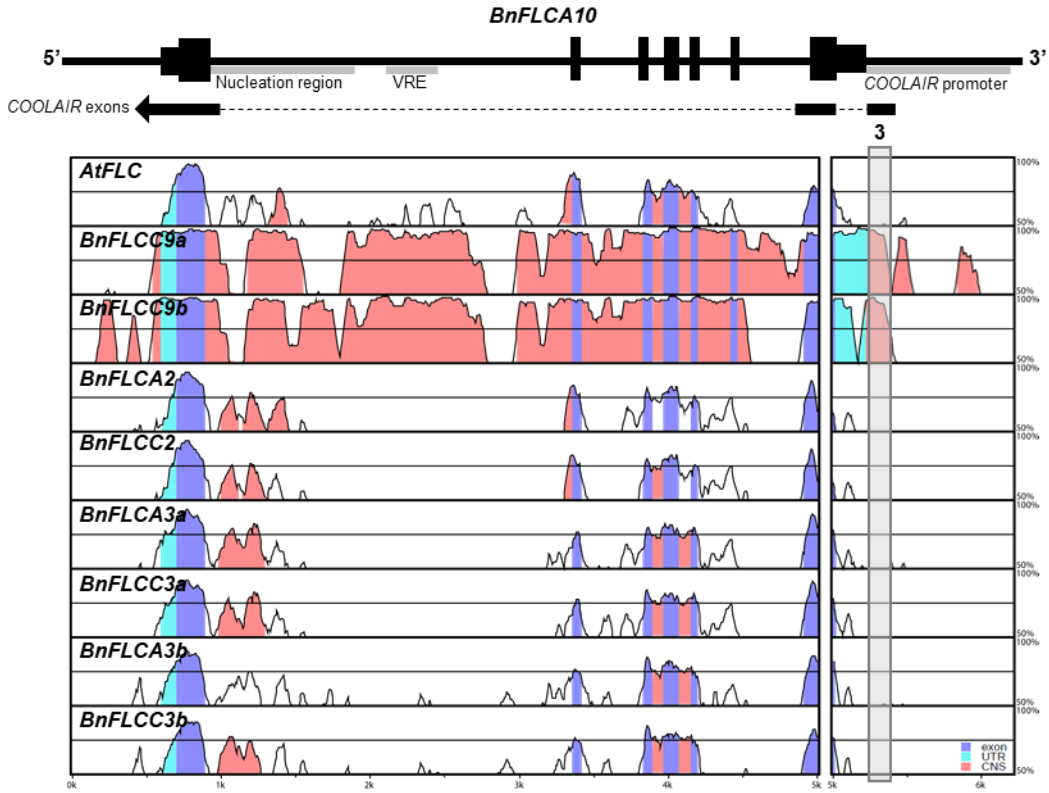


Figure S2: Sequence conservation of Brassica *FLC* homologues against *BnFLCA10*. Sequence conservation of *A. thaliana* and *B. napus* (Cabriolet) A and C genome *FLC* homologues compared with *BnFLCA10*. MLAGAN pair-wise alignments with translated anchoring performed and visualized with the mVISTA online tool, with a window size of 100 bp and cut-off criteria of 70% (Brudno *et al*. 2003; Frazer *et al*., 2004). Coloured areas indicate stretches of homology greater than 75% identity at the nucleotide level. Note the high level of conservation between *BnFLCA10* and its C genome homeologues, *BnFLCC9a* and *BnFLCC9b*. *FLC* sequences were from 0.7 kb upstream of ATG start to 1.2 kb downstream of translational stop, apart from *FLCC3b* which only extended ~0.7 kb downstream of translational stop.

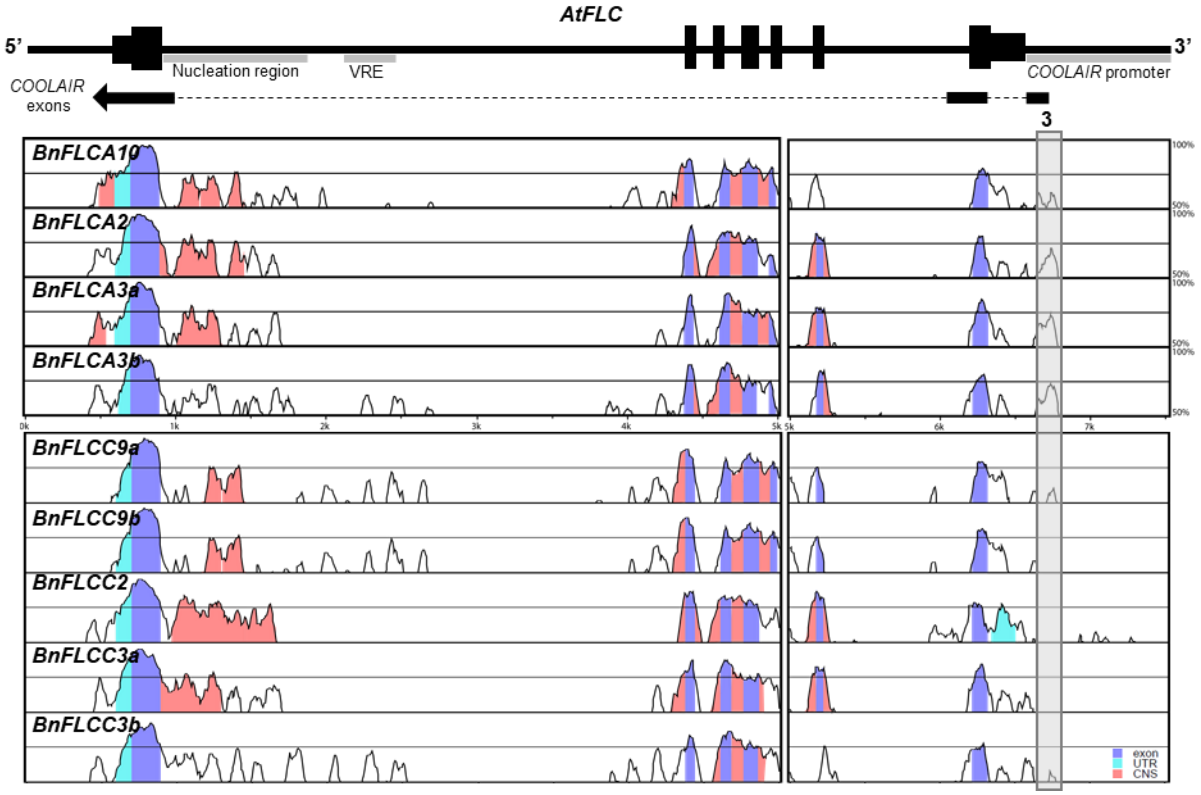


Figure S3: Sequence conservation of Brassica *FLC* homologues against *AtFLC*. Sequence conservation of *B. napus* (Cabriolet) A and C genome *FLC* homologues compared with *AtFLC*. MLAGAN pair-wise alignments with translated anchoring performed and visualized with the mVISTA online tool, with a window size of 100 bp and cut-off criteria of 70% (Brudno *et al*. 2003; Frazer *et al*., 2004). Coloured areas indicate stretches of homology greater than 75% identity at the nucleotide level. *FLC* sequences were from 0.7 kb upstream of ATG start to 1.2 kb downstream of translational stop, apart from *FLCC3b* which only extended ~0.7 kb downstream of translational stop.

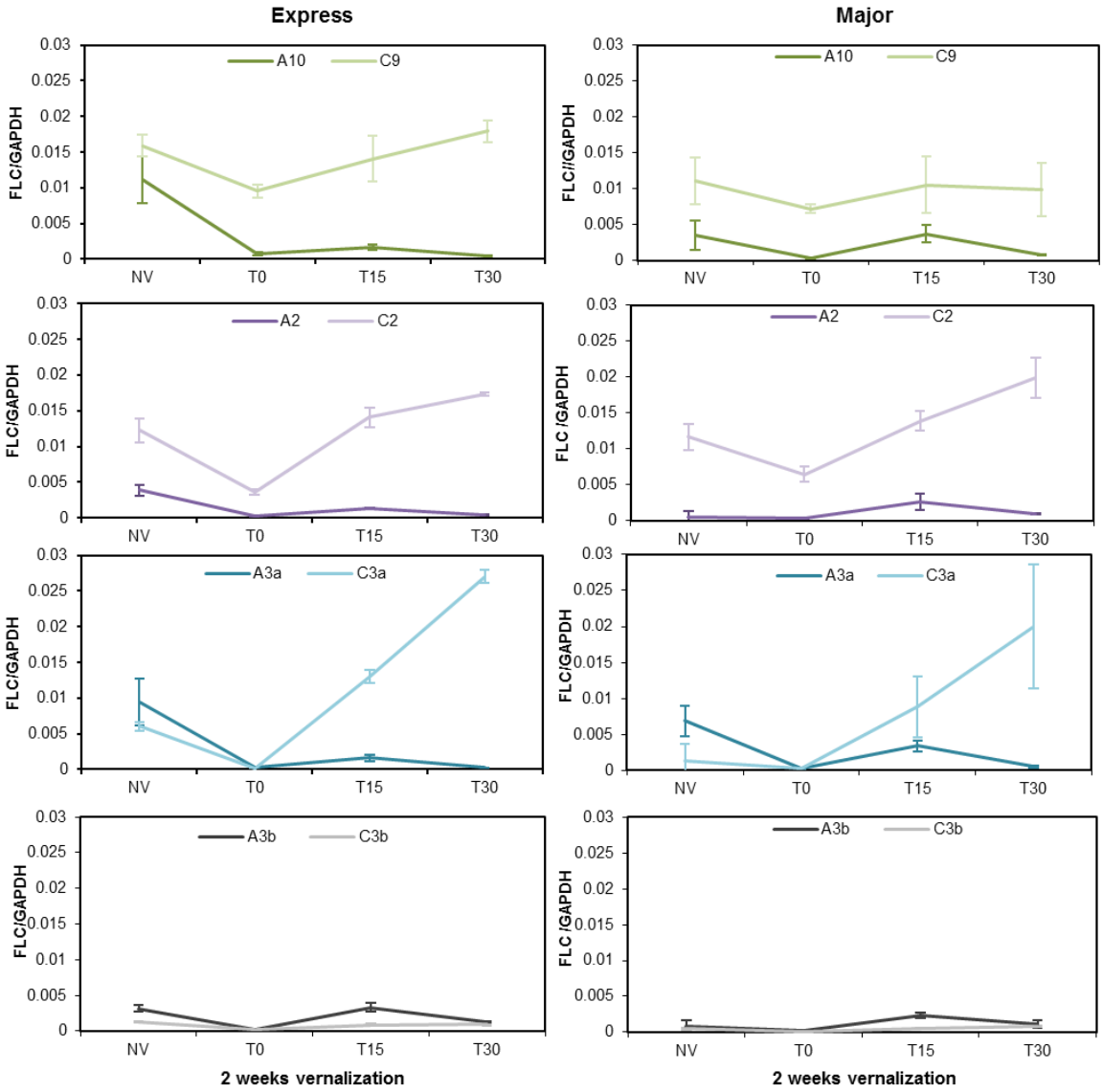


Figure S4: Comparison of *B. napus FLC* homeologue expression patterns after two weeks cold. Expression of spliced sense *FLC* mRNA at NV (non-vernalized) and after two weeks vernalization at three time-points (T0, T15, T30) for *FLC* homeologues in *B. napus* Express and Major winter lines. Samples were normalized to the *GAPDH* reference gene (the *GAPDH** primer set). Error bars are 1 St Dev. from mean for 2-3 biological replicates.

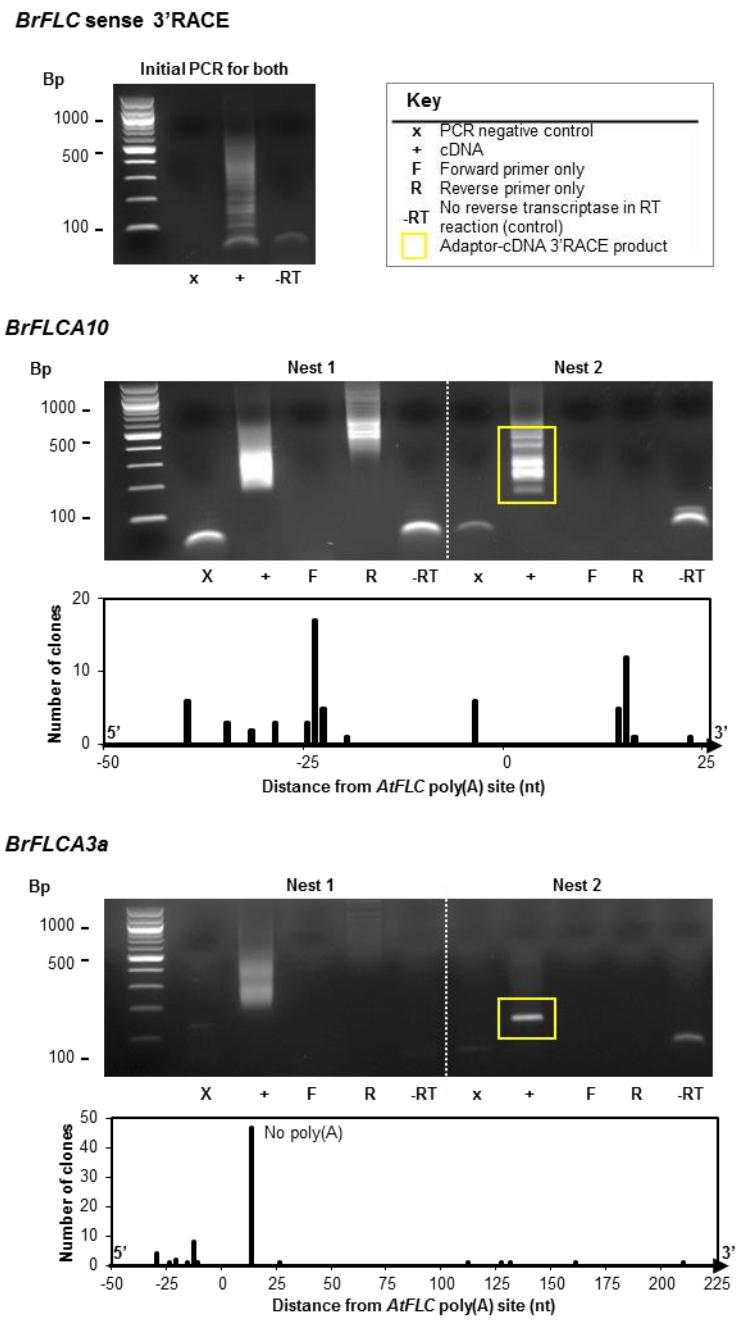


Figure S5: *B. rapa FLCA10* and *FLCA3a* sense transcript polyadenylation sites. 3'RACE experiments probing for *B. rapa* R018 *FLCA10* (top and middle) and *FLCA3a* (top and bottom) polyadenylation sites. Gel images show initial PCR of the cDNA (ligated to an adaptor molecule), which used the same primers for both *FLC* copies and is shown at the top, followed by two rounds of nested PCRs (N1 and N2) with *FLC* specific primers. Graphs detail position of polyadenylation sites relative to the canonical *A. thaliana* proximal or distal site (position 0), and the number of clones (out of up to 96 per experiment) mapped to that site.

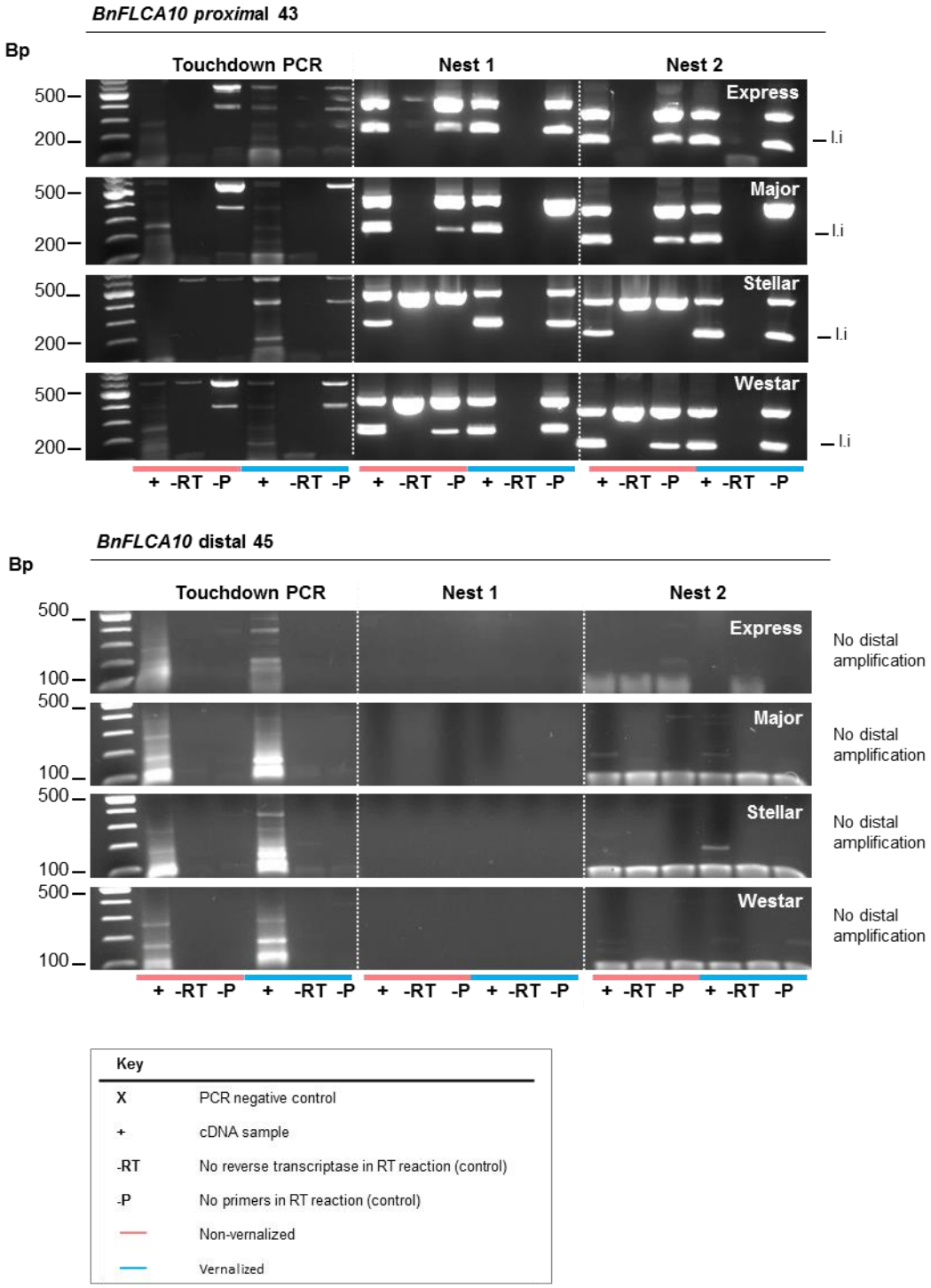


Figure S6: Characterisation of *COOLAIR* in *B. napus FLCA10*. Touchdown nested RT-PCR experiments probed for *FLCA10 COOLAIR* proximal (top) and distal (bottom) transcripts from four *B. napus* cultivars, under non-vernalized (pink line) or two weeks vernalized (blue line) conditions. Initial touchdown PCR was followed by two nested PCRs (Nest 1 and Nest 2), where 43 and 45 refer to the proximal and distal primer sets, respectively. Input cDNA was equalised for the *UBC* reference gene at amplicon level to give semi-quantitative results for the first PCR. Different splice variants have been categorised according to *A. thaliana* classes in Fig. 3.1, for full details see Fig. 4.7.

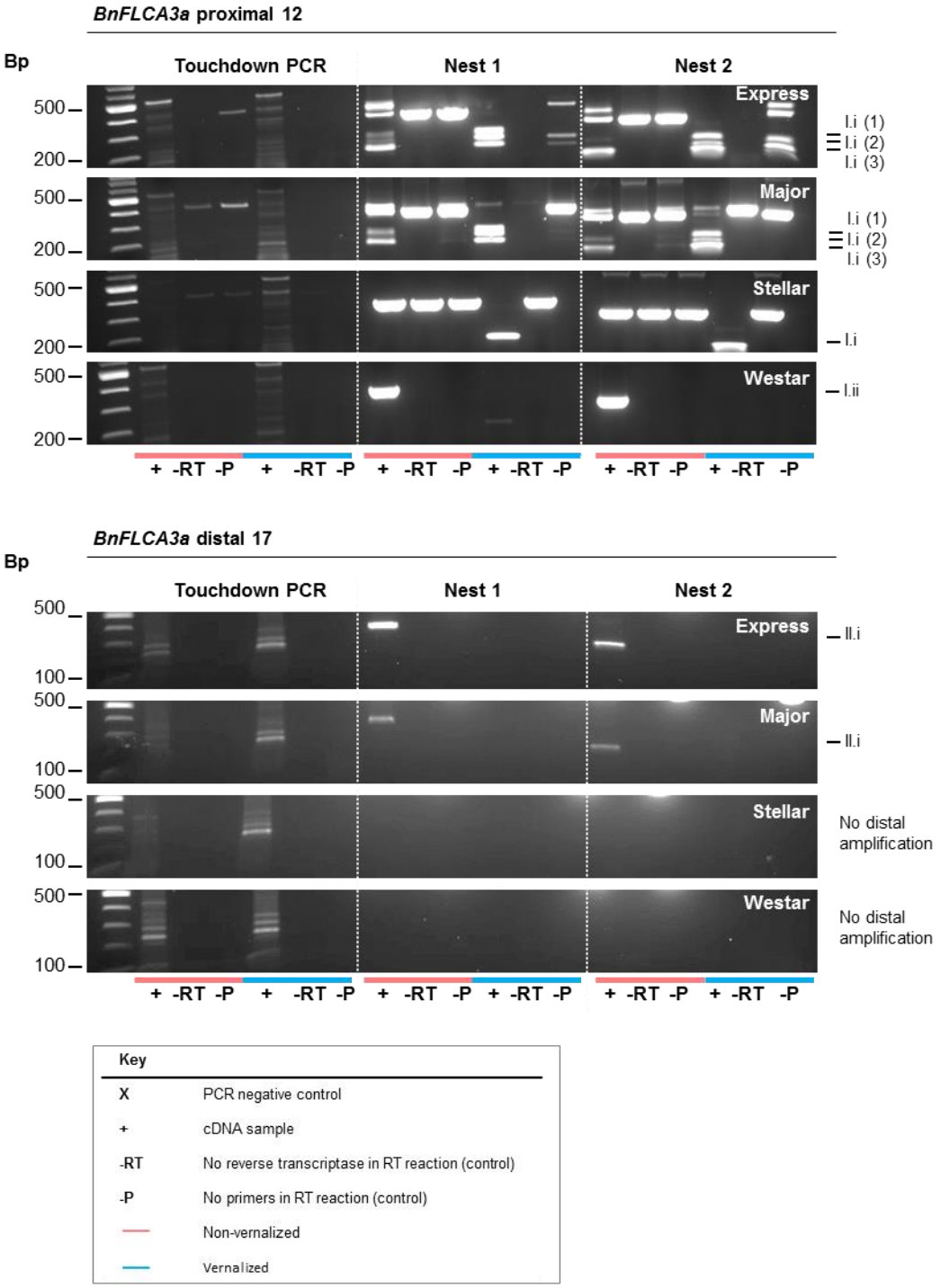


Figure S7: Characterisation of *COOLAIR* in *B. napus FLCA3a*. Touchdown nested RT-PCR experiments probed for *FLCA3a COOLAIR* proximal (top) and distal (bottom) transcripts from four *B. napus* cultivars, under non-vernalized (pink line) or two weeks vernalized (blue line) conditions. Initial touchdown PCR was followed by two nested PCRs (Nest 1 and Nest 2), where 12 and 17 refer to the proximal and distal primer sets, respectively. Input cDNA was equalised for the *UBC* reference gene at amplicon level to give semi-quantitative results for the first PCR. Different splice variants have been categorised according to *A. thaliana* classes in Fig. 3.1, for full details see Fig. 4.7.

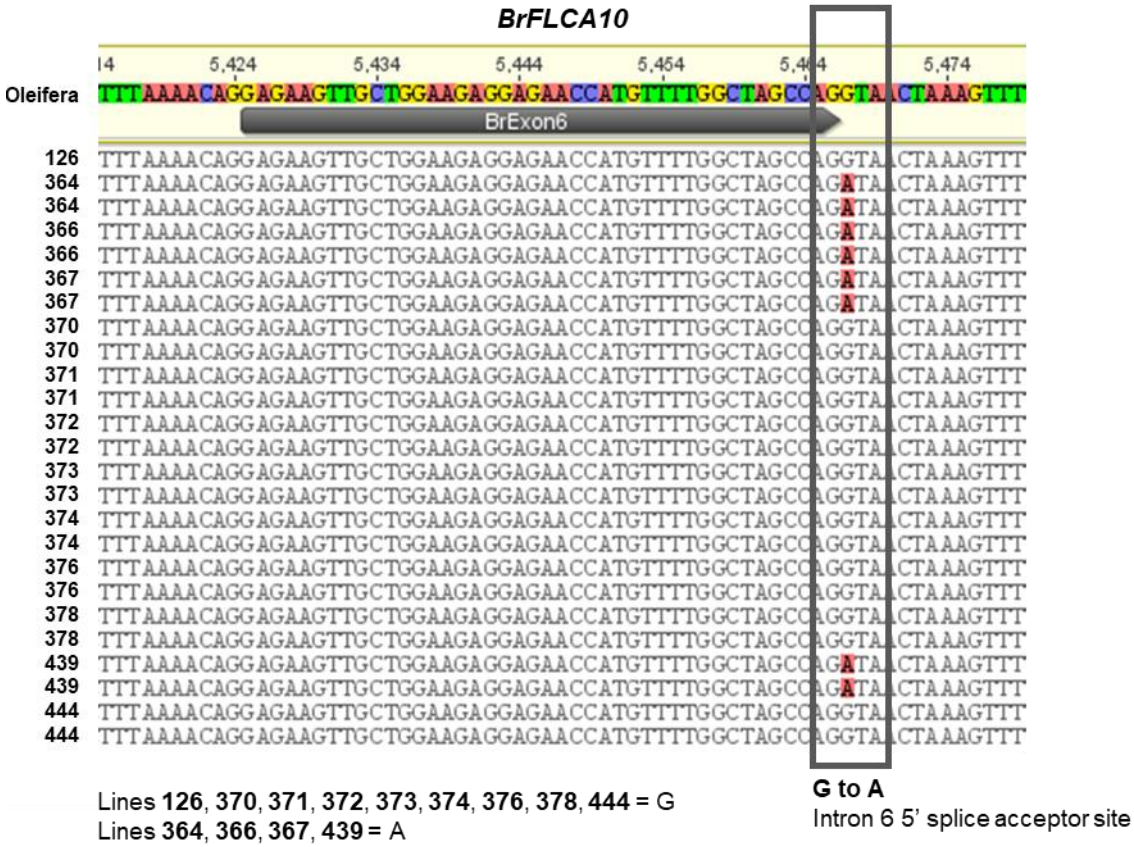


Figure S8: Sequencing of the *FLCA10* exon/inton six boundary in 13 *B. rapa* cultivars. 4/13 lines contained the G to A SNP in intron 6 linked to altered *FLC* sense splicing. Duplicates represent separate PCR reactions and sequencing runs. Lines 374 (Purple Top Milan) and 444 (Maleksberger) were selected for full sequencing reads.

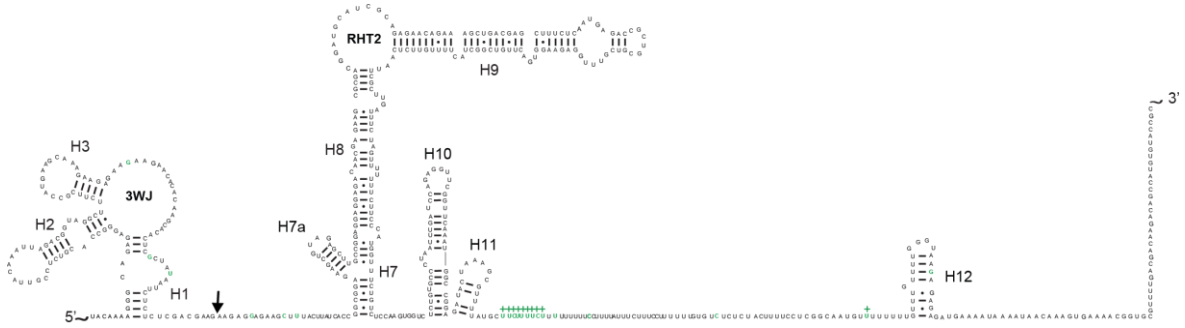


Figure S9: Full-length predicted distal *COOLAIR* secondary structure in the *BrFLCA3a* mutant lines. As in Fig. 5.8, but this shows the full-length sequence. Arrows mark exon-exon boundaries, Nucleotides in green represent SNPs between R018 and PTM *BrFLCA3a* alleles, + represents insertions, and – deletions.

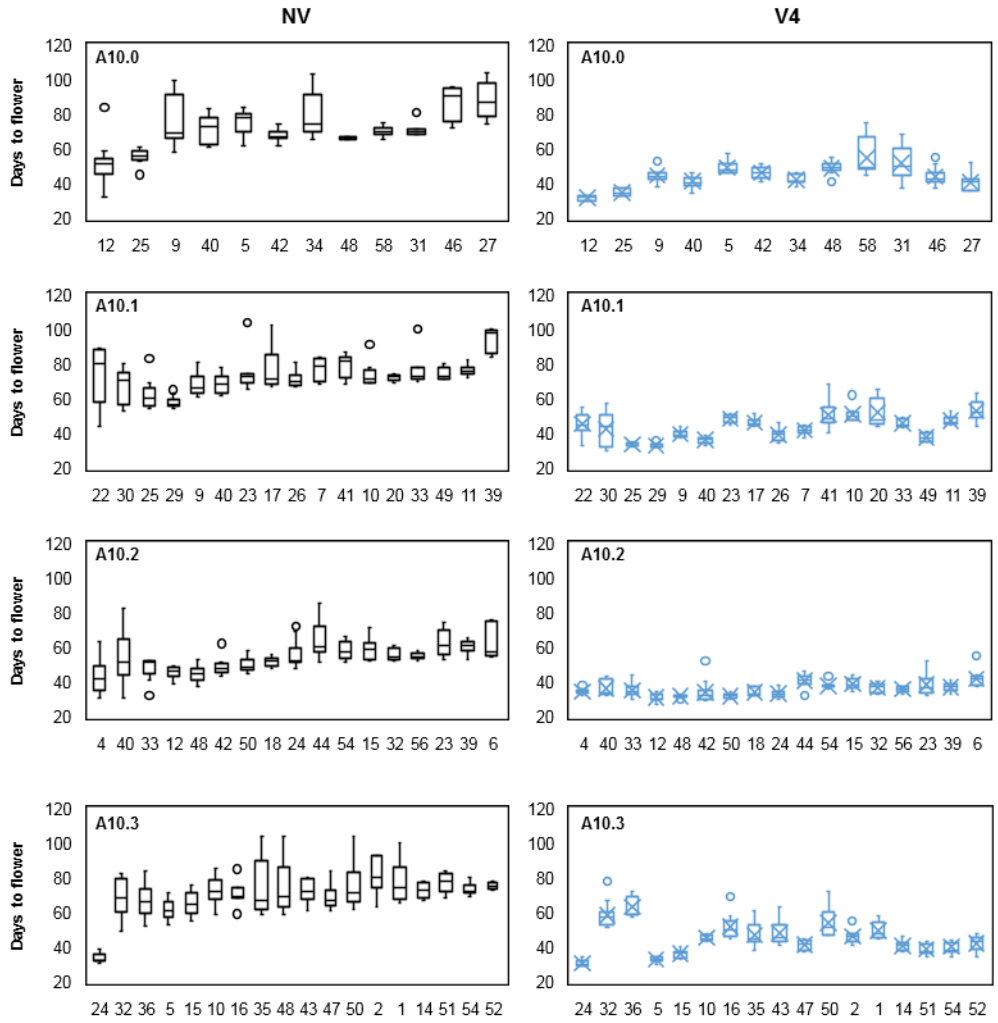


Figure S10:T2 flowering times of *A. thaliana* transgenic plants with wildtype and mutant *BrFLCA10* single copy transgenes. Non-vernalized (left) and four weeks vernalized (V4, right) flowering times for T2 single-copy lines for each *FLCA10* construct, where N represents line number. A10.0 is the wildtype *FLCA10* construct from *B. rapa* PTM; A10.1, A10.2 and A10.3 are mutant constructs.

8.2 SUPPLEMENTARY TABLES

Species	Accession	Homologue	Linker length (nt)
A. alpina	FJ543377.1	*PEP1*	10
A. lyrata	MN47	*FLC1*	13
A. thaliana	Col-0	*FLC*	9
A. thaliana	Var2-6	*FLC*	43
B. oleracea	Alboglabra	*FLCC9a*	7
B. oleracea	Alboglabra	*FLCC2*	45
B. oleracea	Alboglabra	*FLCC3a*	9
B. rapa	R018	*FLCA10*	14
B. rapa	R018	*FLCA2*	46
B. rapa	R018	*FLCA3a*	9
C. rubella	Monte Gargano	*FLC*	19
E. salsugineum	Pall.	*FLC*	7

Table S1: Differences in distal *COOLAIR* linker length across *FLC* homologues. Length of the single-stranded 'linker' region that connects H1 (5' domain) and H4 (central domain) is given in nucleotides for predicted or SHAPE-probed secondary structures (presented elsewhere in this thesis). Note that the 45 nt and 46 nt full-length linker sequences were not shown in the *BoFLCC2* and *BrFLCA2* structure diagrams due to space restrictions.

8.3 SUPPLEMENTARY METHODS

Mini Hot Phenol RNA extraction

Adapted from Box *et al*. (2011).

Materials

Homogenisation Buffer (100 ml):

10 ml 100 mM Tris pH 8-9

1 ml 5 mM EDTA pH 8

2 ml 100 mM NaCl

5 ml 0.5% SDS

72 ml autoclaved H_2O

Safety instructions

Steps 4-17 should be carried out in fume hood.

PPE: nitrile gloves, lab coat and safety glasses.

Method

1. Heat water in plastic beaker in microwave until ~ 60 °C
2. Make up 1% β-mercaptoethanol in Homogenisation Buffer i.e. 100 µl β-mercaptoethanol + 10 ml Homogenisation Buffer.
3. Warm 1% β-mercaptoethanol solution in the beaker of hot water.
4. Add 600 µl of above heated mixture to samples in 1.5 ml tubes at room T. Samples should still be frozen immediately before use (i.e. use straight from liquid N_2 or -80 °C / temporarily store on dry ice).
5. Shake vigorously to mix and then shake vertically in electric mixer for 10 mins.
6. Add 300 µl phenol pH 4.3. Shake vigorously to mix and then shake vertically in electric mixer for 15 mins.
7. Add 300 µl chloroform. Shake vigorously to mix and then shake vertically in electric mixer for 15 mins.
8. Centrifuge at top speed Room T 10 mins.
9. Transfer top aqueous layer to fresh 1.5 ml RNase-free tube (~600 µl).
10. Add 1 x volume (~600 µl) phenol:chloroform. Shake vigorously to mix and then shake vertically in electric mixer for 10 mins.
11. Centrifuge at top speed Room T 10 mins.
    Note: from this step onwards samples should be treated as RNA i.e. use gloves, keep on ice where possible and centrifuge at 4 °C.
12. Transfer the top aqueous layer to fresh 1.5 ml RNAse-free tube (~500 µl).
13. Add 50 µl 3M Sodium Acetate and 450 µl isopropanol (room T). Shake to mix.
14. Incubate at – 80 °C or on dry ice for 15 + mins.

15. Centrifuge at top speed 4 °C 30 mins.
16. Remove supernatant using p1000/p200 pipette, being careful not to disturb pellet. Spin down briefly and remove any remaining supernatant with p200.
17. Place 1.5 ml tubes upside down with lids open on blue roll in fume hood for 10 mins to allow any remaining liquid to evaporate.
18. Re-suspend pellet in 500 µl RNase-free H_2O. Can vortex if do not need DNA.
19. Add 500 µl 4M Lithium Chloride.
20. Incubate at 4 °C O/N.

Next day:

21. Centrifuge at top speed 4 °C 30 mins.
22. Remove supernatant using p1000/p200 pipette, being careful not to disturb pellet.
23. Add 1 ml 80% ethanol. Flick pellet to release from side of tube.
24. Centrifuge at top speed 4 °C 5 mins.
25. Remove supernatant using p1000/p200 pipette, being careful not to disturb pellet. Spin down briefly and remove any remaining supernatant with p200.
26. Incubate 1.5 ml tubes at 50 °C for 10 mins to allow any remaining supernatant to evaporate.
27. Re-suspend pellet in ~50 µl H_2O. Can vortex if do not need DNA.

ExoSAP clean-up of PCR products

1. Prepare ExoSAP mix (10 µl per sample) on ice as follows:
    0.025 µl Exonuclease I (20 U/µl) (New England Biolabs)
    0.25 µl rAPID Alkaline Phosphatase (1 U/µl) (Roche Diagnostics)
    9.725 µl H_2O
2. Add 10 µl ExoSAP mix to ~25 µl PCR product.
3. Heat at 37 °C for 30 min, 95 °C for 5 min, then cool to 4 °C.

Isopropanol precipitation of RNA (to concentrate)

1. Add 0.1 x volume 3M Sodium Acetate and 1 x volume Isopropanol
2. Incubate -20 °C overnight.
3. Centrifuge at max. speed 4 °C 10 minutes.
4. Wash in 75% ethanol
5. Dry pellet and re-suspend in RNase-free water.

CTAB DNA Extraction

Materials

CTAB Extraction Buffer (100 ml, prepare fresh):

10 ml 1M TrisCl pH 7.5

28 ml 5M NaCl

4 ml 0.5M EDTA

58 ml autoclaved H_2O

2g CTAB (dissolve by heating to 65°C)

Wash Solution 1 (100ml):

76 ml EtOH absolute

6.7 ml 3M NaOAc

17.3 ml autoclaved H_2O

Wash Solution 2 (100ml):

76 ml EtOH absolute

1 ml 1M NH_4OAc

23 ml autoclaved H_2O

Safety instructions

Steps 2-13 should be carried out in fume hood.

Note: CTAB is carcinogenic so gloves must be worn when handling.

Method

1. Prepare fresh CTAB Extraction Buffer and preheat to 65°C in a water bath
2. Add 1ml of preheated CTAB Extraction Buffer to 2 ml tube with ground leaf tissue. Vortex until all powder is in suspension
3. Incubate the tube at 65°C for 90 min, mixing every 15 min.
4. Cool samples at room temperature for 10 min.
5. Extract once with 450 µl chloroform:isoamylalcohol (24:1) and mix tubes by inverting for approximately 5 min.
6. Centrifuge at room temperature for 20 min at 3500 rpm.
7. Transfer aqueous phase to a new 2 ml tube.
8. Precipitate the DNA with 900 µl isopropanol and mix well.
9. Centrifuge for 4 min at max. speed and remove supernatant.
10. Wash the pellets with 500 µl Wash Solution 1. Leave for 20 min in the solution.
11. Spin for 2 min at max. speed and remove supernatant.
12. Wash with 500µl Wash Solution 2. Leave for 10 min in the solution.
13. Spin for 2 min at max. speed and remove supernatant.
14. Dry DNA in a heat block at 37 °C (sample tube lids open).
15. Dissolve the DNA in 40 µl 1 x TE Buffer.
16. Shake for 30min at 450 rpm room temperature.
17. Add 1µl of 1U/µl RNAse ONE (Promega) and incubate for 20 min at 37 °C.
18. Store at 4 °C or lower.

Triparental mating

Day 1:

1. Set up o/n 10 ml culture of C58 (pGV2260) from glycerol stock in LB + 200 µg/ml Rifampicin, shake at 28 °C (needs to grow to stationary phase by afternoon of Day 3).

Day 3:

2. Set up 10 ml cultures of HB101 (pRK2013) helper plasmid from glycerol stock in LB + 50 µg/ml Kanamycin, shake at 37 °C until they reach mid-log phase (about 3 hours).
3. At the same time as Step 2, set up 10 ml cultures of SLJ plasmids containing constructs from glycerol stocks in LB + 10 µg/ml tetracycline, shake at 37 °C until they reach mid-log phase (about 3 hours).
4. Then, set up the following in 1.7 ml tubes and briefly vortex:

    Construct:
    0.8 ml C58
    0.2 ml HB101
    0.2 ml construct/LB control

5. Centrifuge cells at 6000 rpm 2 min and remove all supernatant except ~100 µl.
6. Resuspend cells, pipette onto LB plate (without selection) and spread evenly. Incubate at 28 °C o/n (do not turn plates upside down).

Day 4:

There should be an even growth of cells on the plates.

7. Prepare 3 x 1 ml 10 mM $MgSO_4$ in 1.7 ml tubes for each construct and the control (label 100, 10-2, 10-4).
8. Scrape cells (5 - 6 scrapes) off LB plate with loop and add to 100 tube. Vortex well.
9. Add 10 ul of this to 10^{-2} tube and vortex.
10. Add 10 ul of this to 10^{-4} tube and vortex.
11. Pipette 20 ul of each onto a 1/3 sector of a LB + 200 µg/ml Rifampicin + 1 µg/ml tetracycline and spread within each sector.
12. Allow to dry and incubate at 28 °C for 2-3 days - single colonies should be visible in the 10-4 dilution on the construct plates only.

Agrobacterium-mediated floral spray

Method adapted from Clough and Bent (1998).

Safety instructions

For step 4-6 prepare work area by cleaning bench surface with 70% ethanol and then covering with taped down biohazard bags. Dispose of (or autoclave) contaminated equipment.

Method

1. Grow plants for transformation until just bolted and flowering.
2. Innoculate 10 ml of LB (+ 200 µg/ml Rifampicin 1 µg/ml tetracycline) with *Agrobacterium* cells positive for your construct (from single colony or glycerol stock). Shake at 28 °C for 36-48 hours.
3. Add 5 ml above culture to 500 ml LB (+ 200 µg/ml Rifampicin 1 µg/ml tetracycline) and shake at 28 °C for at least 25 hours, until OD_{600} is ~4-4.5.
4. Divide 500 ml culture between two 250 ml plastic bottles and centrifuge at 3500 rpm for 14 min at 15 °C (using the SLC-1500 GSA rotar).
5. Discard supernatant and add a small volume of 5% sucrose, swirl to rinse pellets and discard.
6. Resuspend each pellet in a small volume of 5% sucrose 0.02% Silwet L-77. Dilute to OD_{600} = 1.0 with more 5% sucrose 0.02% Silwet L-77 solution and then pour into a disposable spray bottle.
7. In the glasshouse, place plant trays into individual biohazard bags for spraying.
8. Spray plants in bags thoroughly with the *Agrobacterium* solution. ~500 ml solution is ample amount for 80 plants (2 x trays of 40 plants).
9. Fold plastic bags over and secure ends with tape to keep plants humid overnight.
10. Remove bags the next day and allow to set seed.

Liquid seed sterilization (large-scale)

This method is designed for sterilization of large numbers of seeds in 50 ml falcons.

1. Fill ~10% of 50 ml falcon with threshed seeds.
2. Add 40 ml 70% ethanol and shake 1 min. Pour or pipette off.
3. Add 40 ml 10% bleach 0.5% Triton X and shake 15 min. Pour or pipette off.
4. Move to laminar flow hood and wash seeds at least 3 x with sterile water.
5. After final wash, resuspend in 0.15% agar in sterile water so that seeds are evenly spread.
6. Pour/pipette seeds onto prepared GM-Glu (+ selection) plates, allow to dry and then seal.