Unravelling the floral transition in *Brassica oleracea* using transcriptomics

Shannon Mae Woodhouse

A thesis presented for the degree of
Doctor of Philosophy

University of East Anglia, UK
John Innes Centre, UK
September 2021

Contents

Abstract

Brassica oleracea is an economically important crop species that exhibits extensive morphological diversity. The harvestable product can be vegetative or floral tissues, therefore understanding the genetic basis of the floral transition is an important goal for growers and breeders. Current knowledge of the floral transition largely stems from work in the model species *Arabidopsis thaliana*. Efforts to translate knowledge from this model into *B. oleracea* are complicated by the fact this species is a mesopolyploid and consequently contains multiple paralogues of many genes.

Here we present results from phenotyping a diverse set of 69 *B. oleracea* lines for heading and flowering traits under a range of conditions. We present a novel associative transcriptomics pipeline and use it to analyse the phenotyping results, identifying candidates for vernalisation response in *miR172D* and *BoFLC.C2*. Furthermore, we present a transcriptome time series experiment to investigate gene expression across the floral transition in *B. oleracea* and identify a critical point in floral initiation, at which we see enrichment for gene ontology terms associated with meristem identity. Finally, we adapt an existing method for the comparison of gene expression profiles between *A. thaliana* and *B. oleracea* and find that expression between the two species is largely similar but differently synchronised, but we identify differences in expression profiles between paralogues of key floral genes.

This thesis provides a novel associative transcriptomics pipeline for *B. oleracea* to identify candidates for complex traits. Using flowering time as an example trait, we identify key candidates for the vernalisation response in *B. oleracea.* Furthermore, we provide transcriptome time series data to investigate the floral transition in *B. oleracea* and demonstrate that gene expression dynamics for many genes are similar between *A. thaliana* and *B. oleracea,* but differently synchronised. These results provide a valuable foundation for understanding flowering time in *B. oleracea.*

List of Tables

List of Figures

List of Publications

This thesis includes material from the published works below:

Woodhouse, S., He, Z., Woolfenden, H., Steuernagel, B., Haerty, W., Bancroft, I., Irwin, J.A., Morris, R.J. and Wells, R., 2021. Validation of a novel associative transcriptomics pipeline in *Brassica oleracea*: identifying candidates for vernalisation response. BMC genomics, 22(1), pp.1-13.

Contribution: Designed experiments, alongside JI, RM and RW, carried out all experiments, performed all phenotyping of material, all analyses and produced all figures. Planned and drafted the manuscript, followed by refining with contributions from all authors.

Calderwood, A., Hepworth, J., Woodhouse, S., Bilham, L., Jones, D.M., Tudor, E., Ali, M., Dean, C., Wells, R., Irwin, J.A. and Morris, R.J., 2021. Comparative transcriptomics reveals desynchronisation of gene expression during the floral transition between Arabidopsis and *Brassica rapa* cultivars. Quantitative Plant Biology, 2.

Contribution: Performed experiments and provided feedback to contribute to both the final analysis and manuscript.

Calderwood, A., Lloyd, A., Hepworth, J., Tudor, E.H., Jones, D.M., Woodhouse, S., Bilham, L., Chinoy, C., Williams, K., Corke, F. and Doonan, J.H., 2021. Total *FLC* transcript dynamics from divergent paralogue expression explains flowering diversity in *Brassica napus*. New Phytologist, 229(6), pp.3534-3548.

Contribution: Performed experiments and provided feedback to contribute to both the final analysis and manuscript.

Acknowledgements

I would first like to thank my wonderful supervisory team, past and present, Richard Morris, Rachel Wells, Wilfried Haerty and Judith Irwin. I feel as though I am very fortunate to have had such a supportive and encouraging supervisory team, thank you for giving me the freedom to become a more confident researcher and for making every meeting an enjoyable and helpful discussion. Richard has been a brilliant mentor, whose enthusiasm, dedication to science and love of Bayes, has motivated me through the last four years. Thank you for giving me the opportunity to study for a PhD in your fantastic group (despite my initial lack of computational skills!) and for your guidance and support, you've set an example I can only hope to follow. Thank you to Rachel, for stepping seamlessly into the role of my secondary supervisor. Your ideas and advice have been invaluable and your passion for science and willingness to help others has been a true inspiration. Thank you Wilfried, for being a fountain of ideas and knowledge, your input has really helped me shape my project, thank you for always making time to help me. Finally, I would like to thank Judith for introducing me to the wonderful world of Brassicas and for sharing her knowledge and experience with me, thank you for making me feel welcome and at ease from the start.

I would like to thank all the Postdocs and students in Richard's lab, you have been a wonderful team to work with and have made many lab meetings and social events great fun. I would particularly like to thank Hugh Woolfenden for the enjoyable and helpful discussions and for hints and tips on how to improve my coding, and Ruth Kristianingsih, whose infectious smile and curiosity made her a pleasure to work alongside. I would like to give a special thank you to past lab member Alex Calderwood, thank you for your advice and patience, you helped me more than you know and taught me so much. Thank you for paving the way with your amazing Brassica work, but most of all thank you for your friendship.

Thank you to the brilliant BRAVO team for all their comments and suggestions, being part of such a collaborative group has shown me what can be achieved when scientists come together and share ideas and expertise. I would like to thank Ian Bancroft, Lenka Havlickova and Andrea Harper for their helpful advice and input on my associative transcriptomics work. I would also like to thank Eleri Tudor whose kind words and cups of tea helped me through my first year, thank you for being a brilliant friend and wonderful teacher. Thank you to Jo Hepworth, Catherine Chinoy and Lorelei Bilham for all their help and advice in the lab, I loved working with every one of you.

I would also like to thank the Horticultural team, particularly Lionel Perkins and Damian Alger for their hard work with my Brassica plants, without you I wouldn't have the data for this thesis. I would also like to thank the lovely Dawn Nash for making my tea breaks full of laughter, thank you for always brightening my afternoons.

I was fortunate enough to participate in a great deal of outreach and public engagement during my PhD and for this I would like to thank Carl Harrington for placing his trust in me on many occasions to run outreach activities and for the enthusiasm he brought to every event. I would also like to thank the SAW Trust for having me for my PIPs Internship. I loved every minute of this placement and it only worked to fuel my passion for science communication. The work carried out by Jenni Rant and Sami Stebbings is amazing, and I feel so lucky to have been a part of it.

A special thanks to Alison Smith for welcoming me into her group part-time as a very nervous undergraduate student. This opportunity had a huge impact on me and taught me so many skills. I would also like to thank Marilyn Pike. Marilyn is a brilliant scientist and patient teacher, she believed in me from the start and made me realise that a career in science was a possibility. Marilyn, you were a wonderful mentor and continue to be a wonderful friend. From my time at UEA, I want to thank Colwyn Thomas, whose encouragement and guidance helped me tremendously, thank you for all your support.

Of course, no PhD can be successful without cake, and I would like to thank the lovely Joanna Jennings for making sure my PhD was filled with an abundance of it! Thank you for being a brilliant friend and helping me retain my sanity throughout my PhD, I'm so glad we got to share this journey together. Finally thank you to Elliott. I am so lucky to have you, you've put up with my tears and self-doubt, cheered me on through every achievement and always reminded me what is important. Thank you for your love and encouragement – I couldn't have done it without you.

Shannon Woodhouse
John Innes Centre, Norwich
September 2021

Chapter 1

Introduction

1.1 The origins of *Brassica oleracea* and why flowering time is important

The *Brassicaceae* family exhibits extensive genetic and morphological diversity, including over 3700 known species (Liu *et al.*, 2014). This family contains some of the World's most economically important crops (Table 1.1), with Brassica vegetables being a staple to diets in many parts of the globe (Maggioni, 2015). In 2020, over 84 thousand tonnes of Calabrese (broccoli) alone were produced in the UK (Brown, 2021), indicating just how important the success of these crops is. Within the *Brassicaceae* family, there are six cultivated members. These are explained by the 'Triangle of U', which was determined by cytogenetic analysis and crossing experiments (Nagaharu, 1935). There are three diploid members, *Brassica rapa, Brassica nigra and Brassica oleracea*, which represent the A, B and C genomes respectively and by hybridisation they have produced three allotetraploids, *Brassica juncea* (AABB), *Brassica napus* (AACC) and *Brassica carinata* (BBCC).

Here we focus on the diploid species *B. oleracea*. Many cultivated Brassica vegetables arose from their native wild form *B. oleracea var. oleracea*. Wild cabbage, *B. oleracea* L., is a cruciferous perennial growing naturally along the coastlines of Western Europe (Maggioni *et al.*, 2010). From this single species, selective breeding efforts have enabled the production of the numerous subspecies with diverse morphology. Various parts of Brassicas are harvested, including leaves (e.g. leafy-kale and cabbage), stems (e.g. kohl-rabi), and inflorescences (e.g. broccoli and cauliflower) (Parkin *et al.*, 2014). Such variation means that polymorphism within *B. oleracea* is one of the most prominent examples of crop variability as a direct result of human selection. For all *B. oleracea* subspecies, the shift from the vegetative to the reproductive phase plays an important role in crop production. For leafy types we want to delay this transition, such as in cabbages and kale, whilst for inflorescence types, like broccoli and cauliflower, it is the floral tissue that is harvested, therefore we want the plants to transition, but then hold at a specific point in their development. There are other tissues harvested from different *B. oleracea*, including stalks and auxiliary buds, and for each of these an understanding of the floral transition is critical to their production. As a result of this, being

able to genetically manipulate the vegetative-to-floral transition will aid the development and production of synchronous Brassica vegetables.

Table 1.1: The main Brassica crops and the tissue used for consumption. Table taken from Cartea et al. 2011.

Species	Group	Common name	Organ
Brassica oleracea	*acephala*	Kale, collards	Leaves
	capitata	Cabbage	Terminal leaf buds (heads)
	capitata sabauda	Savoy cabbage	Terminal leaf buds (heads)
	costata	Tronchuda cabbage	Loose heads
	gemmifera	Brussels sprouts	Vegetative buds
	botrytis botrytis	Cauliflower	Inflorescences
	botrytis italica	Broccoli	Inflorescences
	gongylodes	Kohlrabi	Stem
	albogabra	Chinese kale	Leaves
Brassica rapa	*chinensis*	Pak choi, bok choy	Leaves
	dichotoma	Brown sarson, toria	Seeds
	narinosa	Chinese flat cabbage, wutacai	Leaves
	nipposinica	Mibuna, mizuna	Leaves
	oleifera	Turnip rape, rapeseed	Seeds
	pekinensis	Chinese cabbage, pe-tsai	Leaves
	perviridis	Komatsuna, Tendergreen	Leaves
	parachinensis	Choy sum	Leaves
	rapa	Turnip, turnip greens	Roots, leaves and shoots
	ruvo	Broccoleto	Shoots
	trilochularis	Yellow sarson	Seeds
Brassica napus	*pabularia*	Leaf rape, nabicol	Leaves
	napobrassica	Swede, rutabaga	Roots
Brassica juncea	*rugosa*	Mustard greens	Leaves
	capitata	Head mustard	Heads
	crispifolia	Cut leaf mustard	Leaves

The *Brassicaceae* family shares a common ancestor with Arabidopsis ~20 million years ago (Yang *et al.*, 1999). *B. oleracea* exhibits a high level of genomic diversity and this is partly

due to the fact it is a mesopolyploid due to a whole genome triplication event which occurred in an ancestral genome similar to that of *Arabidopsis thaliana* (Arabidopsis), around 23 Mya (Arias *et al.*, 2014). This genome triplication event gave rise to the modern Brassica species (Beilstein *et al.*, 2010). *B. oleracea* also carries remnants from three whole genome duplication events that occurred in its evolutionary history, known as the α, β and γ events (Bowers *et al.*, 2003). As a result of this ploidy, *B. oleracea* often exhibits multiple copies of genes, for example whilst the model species Arabidopsis has only one copy of *FLOWERING LOCUS C* (*FLC*), within *B. oleracea* there are five copies (Irwin *et al.*, 2016). Further to this ancient genome triplication event, comparative analyses between the linkage maps of Arabidopsis and *B. oleracea* have revealed further genome duplications and triplications (Lan *et al.*, 2000; Babula *et al.*, 2003; Lukens *et al.*, 2003). Whilst the presence of multiple gene copies enables the emergence of novel gene functions, by reducing both natural and artificial selective pressure on any one copy, it also impedes efforts to translate knowledge of gene function from model to crop species (Conant and Wolfe, 2008).

1.2 *Arabidopsis thaliana* as a model for flowering time

Model systems have been critical to the progression of biology, and in the 1970s as a result of advances in molecular genetics, Arabidopsis took its place as the model species of choice for many plant biologists (Otto, 2007; Conant and Wolfe, 2008). Arabidopsis is well placed as a model species due to its short generation time, small size, and ease of reproduction through the vast seed numbers produced from self-pollinating flowers. Furthermore, the small genome size, approximately 125 MB, and the publication of this sequence (Kaul *et al.*, 2000) has cemented this species' position as model of choice. These qualities combined and the tools that have subsequently been produced for work in Arabidopsis, mean it has enabled the study of multiple developmental pathways.

The vegetative-to-floral transition is key to the development of all flowering plants. Optimal timing of the floral transition is equally important for wild and domesticated plants, ensuring reproductive success for plants growing in the wild and maximal yields of plants grown as crops. Much of what is currently known about flowering time stems from work in Arabidopsis. Flowering time is a complex process which involves both endogenous and exogenous cues (Song, Irwin and Dean, 2013). Whilst the complexity of the mechanisms and inputs involved in flowering time enable a fine degree of control and adaptive responses at the phenotypic level, it also makes understanding the complex network behind the floral transition challenging.

1.2.1 Flowering pathways

In Arabidopsis there are five main pathways that control the floral transition: ageing, autonomous, hormone, photoperiod, and vernalisation. The ageing pathway ensures Arabidopsis will flower even no floral inductive signals are detected, in an attempt to produce seed before plant death. The ageing pathway is largely regulated by microRNAs (miRNAs) and *SQUAMOUS PROMOTER BINDING PROTEIN LIKE* (*SPL*) transcription factors. miRNAs are small molecules, usually between 18 and 24 nucleotides in length that do not encode proteins and instead act to silence mRNA (Spanudakis and Jackson, 2014; Teotia and Tang, 2015). There are two miRNA families that are crucial to the floral transition in Arabidopsis, these are miR156 and miR172. miR156 is a floral repressor that is expressed in the juvenile phase and decreases in expression as the plant ages (Wu and Poethig, 2006; Wu *et al.*, 2009; Xu *et al.*, 2016), whereas miR172 is a floral promoter whose expression accumulates as the plant ages (Aukerman and Sakai, 2003). The miR56 family targets ten members of the *SPL* family (Wu and Poethig, 2006), required for upregulation of meristem identity genes such as *AP1* and *LFY*. Conversely, miR172 targets six members of the *APETALA2* (*AP2*)-like floral repressors, including *TARGET OF EAT1* (*TOE1*) and *TOE2* (Aukerman and Sakai, 2003; Wu and Poethig, 2006).

In long and short days, mutants of the autonomous pathway are late flowering, but this phenotype can be overcome by vernalisation (Koornneeff, Dellaert and van der Veen, 1982), indicating the genes in this pathway act upstream of the floral repressor *FLC*, promoting flowering by inhibiting *FLC* expression (Michaels and Amasino, 2001; Simpson and Dean, 2002). The autonomous pathway genes encode proteins that promote flowering independent of photoperiod, and these include, *FCA, FLOWERING LOCUS D* (*FLD*), *FLK, FPA, FVE, FY, LUMINIDEPENDENS* (*LD*) and *RELATIVE OF EARLY FLOWERING 6* (*REF6*) (Simpson, 2004; Cheng *et al.*, 2017).

Whilst many plant hormones have been implicated in the hormone pathway, it is largely dependent on a class of plant hormones called gibberellins (GA) (Wilson, Heckman and Somerville, 1992; Davis, 2009). Mutants that block GA biosynthesis (*ga1-3*) exhibit late-flowering, but this phenotype can be rescued with exogenous application of GA. Exogenous GA application is also capable of speeding up flowering time in wild-type plants under short day conditions (Wilson, Heckman and Somerville, 1992). In contrast, in mutants in which GA

signalling is constitutively active, such as *spindly*, flowering is promoted (Jacobsen and Olszewski, 1993).

The photoperiod pathway enables the plant to sense daylength, which acts as a key indicator of favourable flowering conditions for many plant species inducing the expression of photoperiod related genes. The photoperiod pathway achieves this through a close association between light sensing apparatus and the plant's circadian clock. Photoperiod is sensed in leaf tissue and subsequently a signal, dubbed "florigen" is sent through to the shoot apical meristem (Zeevaart, 1976). This signal was later found to be the protein FT. The circadian clock measures the duration of day and night, and is a regulatory network which maintains a consistent oscillating signal in the plant (McClung, 2006). *CONSTANS* (*CO*), which encodes a zinc finger transcription factor, is a key player in the photoperiod pathway. *CO* is downstream of the circadian clock and *CO* mRNA accumulates and degrades in a regular pattern each day (Suárez-López *et al.*, 2001).

For summer annuals, their lifecycle is heavily dependent on the photoperiod pathway. They germinate in spring, flower in summer, and set seed before the winter takes hold. This is often the strategy adopted by plants from warmer climates, however a plant that is a winter annual cannot rely so heavily on photoperiod. Winter annuals germinate in the late summer or autumn, remain in a vegetative phase over the winter period and then flower in the following spring. If these plants were to heavily rely on photoperiod, there is a risk that the autumn may provide day lengths long enough to induce flowering. This could have detrimental effects on the plant, reducing reproductive success due to seed filling occurring through the winter months. As a consequence, for winter annuals, the vernalisation pathway is critical and ensures flowering does not occur until the plant has been exposed to a sufficient period of cold (Song, Irwin and Dean, 2013).

Vernalisation can be defined as the acceleration of flowering following exposure to a long period of cold, typically between one to three months at temperatures between 1 and 10 °C (Simpson and Dean, 2002). Vernalisation is required for the vegetative-to-floral transition to take place in many *B. oleracea* cultivars. This requirement for vernalisation, or lack thereof, determines whether the plant is a winter annual, perennial or biennial or whether it is rapid-cycling or a summer annual (Chouard, 1960). As a consequence, the response of the plant to vernalisation provides quantifiable variation that has been exploited by breeders to develop varieties with more synchronous heading. In the face of a changing climate such information will be critical to develop future breeding strategies.

There are two major genes involved in vernalisation; *FRIGIDA* (*FRI*) (Napp-Zinn, 1961; Clarke and Dean, 1994; Song, Irwin and Dean, 2013) and *FLC* (Sheldon *et al.*, 1999; Michaels and Amasino, 2001). *FRI* is largely responsible for a plant's vernalisation requirement and therefore for conferring a winter growth habit. The FRI protein has a coiled coil domain that directly interacts with the nuclear cap-binding complex and works to promote the accumulation of *FLC* mRNA (Simpson and Dean, 2002; Geraldo *et al.*, 2009). *FRI* activates *FLC* transcription by recruiting chromatin modifiers (Choi *et al.*, 2011), including a histone methyltransferase called *EARLY FLOWERING IN SHORT DAYS* (Kim *et al.*, 2005; Hyun, Noh and Song, 2017). It has been demonstrated that there are two *FRI* copies in *B. oleracea*, one of which has been identified as being functional through complementation assays within Arabidopsis. The genomic location of these copies is not syntenic with Arabidopsis due to a recombination event within the evolutionary history of *B. oleracea* (Irwin *et al.*, 2012).

FLC encodes a MADS-box transcription factor that acts as a repressor of flowering. A quantitative relationship has been described between expression levels of *FLC* and flowering time. *FLC* suppresses flowering by binding to the first intron of *FLOWERING LOCUS T* (*FT*), repressing the expression of *FT*. It is also able to suppress other photoperiod pathway genes, *AGAMOUS-LIKE20/SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) and *FD* by direct binding prior to vernalisation (Searle *et al.*, 2006). Exposure to a period of vernalisation that meets the requirements of the plant, decreases the expression of *FLC* and this repression is mitotically stable. *FLC* expression consequently acts as a memory of whether the plant has experienced winter (Song, Irwin and Dean, 2013).

1.2.2 Key Floral Integrators

A core network of genes integrates the above pathways, to enable flowering at the optimal time. Whilst more than 100 floral integrator genes have been identified within Arabidopsis, it is these core integrators and the regulatory links between them, that are crucial to understanding the transition.

Both the photoperiod and vernalisation pathways converge on the floral activator *FT*. The FT protein is a mobile signal and is transported from the leaves to the meristem, where its expression helps to trigger flowering. This is true for many species, not just Arabidopsis, demonstrating that this regulator has been conserved in higher plants (Jaeger *et al.*, 2013). *FT* was identified from a photoperiod sensitive mutant, that exhibited delayed flowering when in short days and this photoperiod sensitivity can be attributed to the direct regulation of *FT* by

CO (Putterill *et al.*, 1995; Samach *et al.*, 2000; Suárez-López *et al.*, 2001). *CO* directly activates *FT* expression by encoding a transcription factor with two B-box zinc fingers (Boss *et al.*, no date). *CO* is a circadian clock gene and it's transcription is regulated through *GIGANTEA* (*GI*) and *CYCLING DOF FACTOR1* (*CDF1*) (Suárez-López *et al.*, 2001). Further to being controlled by the circadian clock, *CO* is directly impacted by day length. Under long day photoperiods, *CO* mRNA will peak at the beginning and end of the photoperiod, but under short days this peak will occur in darkness (Suárez-López *et al.*, 2001).

LEAFY (*LFY*) has been identified as another core flowering gene through work in Arabidopsis and acts in parallel to *FT* (Kardailsky *et al.*, 1999). *LFY* was identified and subsequently named, in a mutant screen as a mutant that produced leafy shoots instead of flowers. Any flowers that were produced in these mutants lacked petals and stamens (Shannon and Meeks-Wagner, 1991). *LFY* plays a role in the floral transition and is also responsible for floral meristem identity in Arabidopsis (Weigel *et al.*, 1992). Working with *FT, LFY* integrates flowering signals to activate floral meristem identity genes (Kardailsky *et al.*, 1999; Kobayashi *et al.*, 1999), such as *APETALA 1* (*AP1*) (Bowman *et al.*, 1993; Ruiz-García *et al.*, 1997; Abe *et al.*, 2005), *AGAMOUS* (*AG*) (Mizukami and Ma, 1997; Busch, Bomblies and Weigel, 1999), *CAULIFLOWER* (*CAL*) (Parcy, Bomblies and Weigel, 2002; Serrano-Mislata *et al.*, 2017) and *TERMINAL FLOWER1* (*TFL1*) (Shannon and Meeks-Wagner, 1991; Serrano-Mislata *et al.*, 2017). *LFY* itself is regulated by *AGAMOUS-LIKE24* (*AGL24*) and *SOC1* (Yoo *et al.*, 2005; Lee *et al.*, 2008). It has been suggested that many environmental pathways converge onto *LFY*, due to its interactions with the photoperiod and hormone pathways (Blazquez *et al.*, 1997; Eriksson *et al.*, 2006), reinforcing its role as a floral integrator.

SOC1 encodes a MADS box transcription factor and integrates signals from the ageing, hormone, photoperiod and vernalisation pathways (Moon *et al.*, 2003; Lee and Lee, 2010). Two antagonistic regulators control *SOC1*, *CO* and *FLC*. *CO* activates *SOC1* through *FT*, whereas *FLC* represses it by directly binding to the promoter site (Lee and Lee, 2010). *SOC1* regulates floral patterning, as well as flowering time, preventing premature differentiation of the meristem in Arabidopsis (Liu *et al.*, 2009; Lee and Lee, 2010). The regulation of *SOC1* is directly tied to another flowering gene, *AGL24*. These two genes regulate each other in a positive feedback loop (Liu *et al.*, 2008) and directly bind to one another, leading to mutual upregulation during the floral transition. Despite their synchronous expression, *SOC1* and *AGL24* affect different downstream targets. *SOC1* binds directly to *LFY* but *AGL24* does not (Liu *et al.*, 2008). This may be attributed to the role of *AGL24* in maintaining inflorescence

identity, rather than floral fate, and is consequently repressed by *LFY* and *AP1* (Yu *et al.*, 2004).

AP1 is another MADS-box containing transcription factor, which plays a role in floral meristem identity and floral organ specification (Alejandra Mandel *et al.*, 1992). Overexpression of *AP1* leads to the conversion of both apical and lateral shoots into flowers (Mandel and Yanofsky, 1995), whilst null mutations in the *AP1* genes result in plants that lack petals (Koornneeff, Dellaert and van der Veen, 1982; Ruiz-García *et al.*, 1997). Evidence suggests that *AP1* directly represses the expression of *AGL24, SOC1* AND *SHORT VEGETATIVE PHASE* (*SVP*). It is known that *AGL24* plays a role in maintaining inflorescence meristems (Yu *et al.*, 2004), whilst *SVP* is conversely responsible for maintaining vegetative meristems (Lee and Lee, 2010), hence the role of *AP1* as responsible for conferring the floral state to the meristem. *AP1* expression is promoted by the interaction between *FT* and *FD* at the apex (Abe *et al.*, 2005; Wigge *et al.*, 2005).

Floral repressors also play important roles in the floral transition, and one such gene is *TFL1* (Shannon and Meeks-Wagner, 1991). *TFL1* is in the same gene family as *FT* but acts antagonistically to it (Serrano-Mislata *et al.*, 2017). Whilst neither of these genes are transcription factors, as they do not possess any DNA binding activity, they both interact with the FD protein, which is a bZIP transcription factor (Abe *et al.*, 2005; Wigge *et al.*, 2005; Jaeger *et al.*, 2013). *TFL1* works to limit the activity of the AP1 and LFY proteins, in order to maintain inflorescence identity (Ratcliffe, Bradley and Coen, 1999). For wild-type Arabidopsis, their flowers develop indeterminately (Shannon and Meeks-Wagner, 1991). The transition to the floral phase leads to the conversion of the vegetative meristem to an inflorescence meristem, which creates the subsequent floral structures, however the shoot apical meristem remains an inflorescence meristem. In *TFL1* mutants, the primary inflorescence meristem converts into a floral meristem (Alvarez *et al.*, 1992), which terminates in a flower, rather than maintaining this indeterminate state we see in wild-type Arabidopsis.

Another key floral repressor is *SVP*. This gene has a dual function, it works as a floral repressor in early development, and as a floral meristem identity gene later in development (Gregis *et al.*, 2013). When acting as a floral repressor, *SVP* forms a heterodimer with *FLC*. *SVP* is known to associate to the promoter regions of both *FT* and *SOC1* to delay flowering (Li *et al.*, 2008). Later in development *SVP* acts redundantly with *AGL24, AP1* and *SOC1* to maintain an indeterminate meristem (Gregis *et al.*, 2008, 2009).

The control logic behind the floral transition is complex and highly interwoven, involving environmental and endogenous factors. The key floral integrators and their combined interactions allow integration of these signals, enabling the floral transition to be robust. Many of the flowering time genes identified in Arabidopsis have been found to be conserved in a range of species (Jung and Müller, 2009; Blümel, Dally and Jung, 2015), but in polyploid crops, such as *B. oleracea*, transfer of knowledge from model to crop is not straight forward.

1.3 Modelling flowering time in Arabidopsis

The gene network that regulates flowering time is a complex system involving many transcription factors that control floral development both temporally and spatially. As we started to understand more about the floral transition, gene regulatory networks began to be described to try and model the transition. One of the earliest approaches to modelling flowering time was developed over 40 years ago (Thornley, 1972). This model uses a reductionist approach, based on the biochemical interaction between two enzymes that catalyse a substrate into two morphogens, of which the concentration leads to a switch between the vegetative and floral states. This model is parameter dependant because the two stable states, vegetative and floral, become one with a change in parameter value. Although this work is applied to flowering time, the simplicity of the model means it could be used to describe many systems with different developmental pathways.

Another model of flowering time within Arabidopsis, which is of particular interest here, was produced by Jaeger *et al.* (2013) (Jaeger *et al.*, 2013). This model is a simplified network which accounts for the major dynamic properties of the floral transition by taking just five floral integrators, *AP1, LFY, FD, FT* and *TFL1*, which are used as the nodes of the network. The edges of the network, which represent regulatory interactions between the nodes, were deciphered both molecularly and genetically. Arabidopsis single and double mutants of the floral integrators were taken, and flowering time was measured using the number of rosette and cauline leaves present at flowering, these measures were used to parameterise the network. The model described a number of key dynamics within the floral transition, including that relative *FT* and *TFL1* levels were critical for determining when the floral transition would occur (Jaeger *et al.*, 2013). This model was extended by Valentim *et al.* (2015) (Valentim *et al.*, 2015), through the use of expression data to further parameterise the model to incorporate additional genes. This meant that the network rather than being focused on gene hubs, now centred around the specific functions of genes themselves.

Whilst the models discussed so far have aimed to describe the key interactions that trigger the vegetative-to-floral transition in Arabidopsis, other models have been developed which are more elaborate and aim to provide a full view of all of the elements involved in flowering in Arabidopsis. Blümel *et al.* (2015) (Blümel, Dally and Jung, 2015) aggregated all known information on the floral transition in Arabidopsis, including mutant analysis, sequencing and complementation analysis or heterologous expression, RNA interference, and clear linkage with a major QTL to provide an in-depth view of all the interactions that combine to produce the floral transition.

1.4 Current knowledge of flowering in Brassicas

1.4.1 The vernalisation pathway in Brassicas

For many *B. oleracea,* a period of vernalisation is needed to induce the floral transition. Here in the UK, this puts our *B.* oleracea production at risk as we are experiencing warmer weather and more variation in our winter temperatures. In the UK, the five warmest winters since 1884, have all occurred in the last 15 years (Kendon *et al.*, 2021). As a result of this, one of the most well studied areas of the floral transition in *B. oleracea* is the vernalisation pathway.

Association studies that focused on mapping the vernalisation response in *B. oleracea* have identified regions containing homologues of *FLC* and *FRI* as candidates for explaining the variation in flowering time (Axelsson, Shavorskaya and Lagercrantz, 2001; Okazaki *et al.*, 2007; Ridge *et al.*, 2015). Both *FRI* and *FLC* orthologues have been identified and characterised and a complex picture is beginning to emerge to describe this pathway in *B. oleracea*. Two orthologues of *FRI* have been identified and characterised, *BolC.FRI.a* and *BolC.FRI.b.* These two genes demonstrate amino acid conservation in the C-terminus and central regions to the Arabidopsis *FRI,* however due to a recombination event, their genomic location is not syntenic with Arabidopsis. *BolC.FRI.a* has been demonstrated to be functional through complementation experiments in Arabidopsis (Irwin *et al.*, 2012), however the structure of the homologues has diverged, with a change in the number of coiled-coil domains present, that could impact protein-protein interactions (Irwin *et al.*, 2012).

B. oleracea contains five *FLC* homologs, *BolFLC.C02, BoFLC.C03a, BoFLC.C03b, BoFLC.C09a* and *BoFLC.C09b* (Lin *et al.*, 2005; Okazaki *et al.*, 2007; Golicz *et al.*, 2016; Irwin *et al.*, 2016; Calderwood, Lloyd, *et al.*, 2021) and these have been located to their

respective linkage groups using allelic sequence variation in lines from a doubled haploid population (Razi *et al.*, 2008). Such a large number of paralogues indicates retention of *FLC* copies was high and this is likely due to their role as transcription factors in regulatory functions key to plant development (Hong *et al.*, 2011). Comparison of the promoter and intragenic regions of Arabidopsis *FLC* to the *BoFLC* homologues has revealed high conservation within segments known to be essential to the regulation of *FLC* expression, however large differences were seen in the overall structure and organisation of the genes (Razi *et al.*, 2008). The greatest differences occur within *BoFLC.C03b* and *BoFLC.C09a,* which are non-syntenic to Arabidopsis *FLC* and are thought to have arisen from a tandem or segmental duplication (Cai *et al.*, 2014).

BoFLC.C2 is by far the most extensively studied *FLC* copy of *B. oleracea* and that is in part due to the large role it plays in determining heading date in *B. oleracea* (Lin *et al.*, 2005; Ridge *et al.*, 2015; Irwin *et al.*, 2016). *BoFLC.C2* exhibits similar genomic organisation to Arabidopsis *FLC*, with seven exons of a similar size (Lin *et al.*, 2005). A study within cauliflower (*B. oleracea var botrytis* L.) using a segregating F2 population demonstrated that *BoFLC.C2* expression accounted for 65% of the variation seen in flowering time within the population, reinforcing the idea that this copy is critical in determining heading date (Ridge *et al.*, 2015). It has also been determined that *BoFLC.C2* expression is diminished in both the leaf and apex tissues during cold treatment, further indicating a role in vernalisation response (Lin *et al.*, 2005). Furthermore, analysis of two commercial *B. oleracea* lines revealed that *cis* polymorphism at *BoFLC.C2* quantitatively influences the degree of cold-induced epigenetic silencing within *B. oleracea* (Irwin *et al.*, 2016). Two *B. oleracea* genotypes, each harbouring different alleles of *BoFLC.C2* were compared in field trials and a clear difference in vernalisation requirement between the two genotypes was identified. Complementation experiments were carried out within Arabidopsis and determined that the allelic variation was a functional consequence for the differences seen in heading date. It was concluded that the allelic variation resulted from *cis* polymorphism (Irwin *et al.*, 2016).

A recent study conducted in *B. napus,* which has nine *FLC* copies, revealed that these copies taken together are responsible for determining vernalisation requirement between accessions (Calderwood, Lloyd, *et al.*, 2021). The relaxed selection pressure, due to the presence of multiple FLC paralogues, has resulted in divergence between the paralogues and variation in *FLC* expression between both paralogues and accessions. This phenomenon has enabled the development of a large range of vernalisation requirements and a significant range of behaviours in response to cold in *B. napus*. It is possible due to the close relationship between *B. napus* and *B. oleracea,* that this is also the case for *B. oleracea.*

1.4.2 Exploring the roles of homologues of Arabidopsis flowering time genes in Brassica

Much work on the floral transition in *B. oleracea* has been carried out on cauliflower and broccoli, due to the range of vernalisation requirements and because of their need to develop inflorescence tissue for harvest. High ambient temperatures can greatly delay harvest time, therefore understanding the genetic regulation of the floral transition in these subspecies, in which inflorescence tissue is the harvestable product, is highly important. One study used GWAS to identify candidates for temperature related floral induction in cauliflower. This analysis identified QTL co-localised with *BoFLC.C2, BoVRN1, BoFLD* and *BoVIN3,* all homologues of Arabidopsis flowering time genes. Transcriptional profiling was also conducted for *BoFLC.C2* and *BoVRN2*. *BoVRN2* levels were found to be higher in faster curding genotypes indicating *BoVRN2* plays a key role in heading (Matschegewski *et al.*, 2015).

Using an F2 population of rapid cycling *B. oleracea,* derived from an F1 var *alboglabara* x var *italica* cross, *BoCO* was found to be associated with changes in flowering time through QTL analysis. The homologue of *BoCO,* Arabidopsis *CO* is a known circadian gene involved in the photoperiod pathway (Bohuon *et al.*, 1998). Another study conducted in cauliflower and broccoli looked at expression profiles of homologues of Arabidopsis MADS-box genes, *BoAP1.a, BoAP1.b, BoCAL, BoFUL.a, BoFUL.b, BoFUL.c, BoFUL.d* and non-MADS-box genes *BoLFY* and *BoTFL1*. This study demonstrated that *BoFUL* and *BoLFY* transcript abundance was the highest the arrest stage of cauliflower, indicating these homologs may have consistent roles to their Arabidopsis counterparts in initiating inflorescence meristems. In contrast, expression of *BoTFL1*, whose ortholog Arabidopsis *TFL1* is a strong floral repressor, did not correlate with the suppression of floral primordia within cauliflower suggesting a different role for this homolog (Duclos and Björkman, 2008). Finally the study demonstrated that maintenance of curd arrest was a consequence of *BoCAL, BoAP1.a* or *BoLFY,* indicating these three genes are all performing roles in inflorescence meristem maintenance consistent with their Arabidopsis orthologs (Duclos and Björkman, 2008).

Despite the numerous studies that have concluded that many *B. oleracea* orthologues of Arabidopsis flowering time genes play similar roles, an in-depth analysis of how the different gene paralogues are behaving is lacking in the literature. Very few *B. oleracea* flowering time

genes have been dissected in as much as the *FRI* and *FLC* gene families, in which it has been demonstrated that individual gene copies have diverged in both behaviour and function.

1.5 Knowledge transfer between Arabidopsis and *B. oleracea*

Whilst we know much about flowering in Arabidopsis and have developed models to describe the interactions which lead to the transition, we still possess relatively limited knowledge on this phenomenon in *B. oleracea*. Despite the close evolutionary relationship between Arabidopsis and *B. oleracea,* a large part of the challenge of transferring this knowledge arises from the genome multiplication events that have occurred in the evolution of *B. oleracea*. The presence of multiple gene paralogues reduces the selective pressure on any one copy, which enables mutations to occur with limited phenotypic effects. Over time these mutations can accumulate and lead to genes acquiring novel functions (neofunctionalisation), losing a subset of their original function (subfunctionalisation) or becoming entirely non-functional (Conant and Wolfe, 2008). As a consequence of this, when transferring knowledge from Arabidopsis to *B. oleracea*, it is important to determine to what extent gene copies have diverged and whether they are performing the same function as the orthologous gene.

In addition to the presence of multiple gene copies, there has been a high level of chromosomal rearrangements in the lineages resulting in Arabidopsis and Brassica and again this is thought to be attributed to polyploidisation (Lukens *et al.*, 2003). Extensive genome duplications have been identified within the Brassica species during early comparative studies using genetic linkage maps (Lagercrantz and Lydiate, 1996). Despite this, there remains substantial collinearity between Arabidopsis and Brassica, which proves advantageous in the transfer of knowledge across species (Parkin *et al.*, 2014). Furthermore, physical mapping approaches in comparative studies of genome microstructure within targeted regions of *B. oleracea, B. napus* and *B. rapa* revealed three paralogous segments in each, consistent with the triplicated nature of the family. Synteny within these regions was found to be well conserved (O'Neill and Bancroft, 2000; Rana *et al.*, 2004; Park *et al.*, 2005; Town *et al.*, 2006), but evidence of extensive gene loss was also identified, which could again pose an issue to knowledge transfer.

An added layer of complexity is introduced when we consider gene regulatory networks. The ploidy present in *B. oleracea* means that not only is a specific transcription factor present in multiple copies, so are its targets and regulators which vastly increases the number of regulatory links within the network. An important step in tackling this is being able to

determine which gene copies are performing what functions. Identifying which paralogues have become redundant and which have experienced neo- or subfunctionalisation is an important step in beginning to simplify the problem. Identifying which genes have retained their function will be a critical step in using current Arabidopsis knowledge to aid the construction of gene regulatory networks in *B. oleracea*.

1.6 Thesis overview

This thesis will investigate flowering time in *B. oleracea* on a genome wide scale, in order to identify key gene candidates and to begin to uncover the gene regulatory network behind the floral transition in this species. This information will provide a good foundation for breeding efforts to future-proof our Brassica crops against ever changing climatic conditions.

Within the next chapter, a large-scale phenotyping experiment is described, in which flowering trait data is collected on a previously unphenotyped panel of diverse *B. oleracea* accessions to assess the effects of vernalisation timing, duration and temperature. This work identifies *B. oleracea* accessions that require minimal or no period of vernalisation to flower. Using this phenotyping data, a novel associative transcriptomics pipeline is produced and validated, and subsequently used to identify candidates for vernalisation response in *B. oleracea* in *BoFLC.C2* and *miRNA172D,* gene candidates that could prove to be useful in future targeted breeding efforts.

The third chapter presents the first study in *B. oleracea* to follow the transcriptome through the floral transition. Transcriptome time series data from leaf and apex tissues of the rapid cycling *B. oleracea* line DH1012 were generated, to explore the gene expression dynamics of homologues of key floral regulators that have been identified within the model species Arabidopsis. Differential gene expression analysis is used to identify differences in the leaf and apex of DH1012 across the floral transition, whilst gene ontology analysis confirms enrichment of the flowering development term. Using differential gene expression analysis, a critical point in development is presented, that mirrors previous findings in the model species Arabidopsis.

Much of the knowledge of flowering time has been generated through studies of the model species Arabidopsis, therefore the fourth chapter explores the transfer of knowledge between Arabidopsis and *B. oleracea.* A gene registration technique developed in *B. rapa* is adapted

for use in *B. oleracea* and subsequently employed to explore the differences in gene expression between Arabidopsis and *B. oleracea.* In depth analysis of the results discovered gene paralogues exhibiting different expression patterns to their homologous Arabidopsis genes. This difference in expression profile suggests these paralogues may have neo- or subfunctionalised and this is explored for paralogues of *AP1* using causal structural inference.

Chapter 2

Phenotyping and associative transcriptomics for Brassica vegetables

2.1 Introduction

2.1.1 Developmental transitions are essential to the successful growth of all plants

Plants undergo several developmental transitions throughout their lifecycle. Before a plant is able to flower, it must first pass through a juvenile phase and into an adult vegetative phase. Only in the adult phase is the shoot apical meristem able to respond to floral inductive signals and subsequently transition to the floral phase of development (Scotr Poethig, 1990; Bäurle and Dean, 2006; Huijser and Schmid, 2011). There are genetic programs which underlie these developmental transitions, and these are sensitive to both environmental and endogenous stimuli. The timing of developmental transitions is critical to the success of the plant and for crop species such as *B. oleracea,* the timing of these transitions is of considerable agronomic importance. For *B.* oleracea, the range in morphology means that for some crop types, such as kale, we do not want the floral transition to take place, but for other crop types, such as broccoli, it is the inflorescence tissue that is harvested, so we want the transition to occur but the plants to hold at a specific point in their development. For all types understanding the control logic behind these developmental transitions is of great importance. If a crop flowers prematurely this can result in a reduction in biomass and seed set (Huijser and Schmid, 2011) and if the development of the plants is not synchronous it can lead to difficulties for growers in knowing when to harvest to achieve the optimal yield. A clear agronomic goal is therefore to produce *B. oleracea* crops with more predictable, synchronised heading dates. To manipulate flowering time in a controlled manner, an understanding of the genetic regulation of developmental transitions and their interactions is essential.

2.1.2 The juvenile-to-adult transition is a precursor to the vegetative-to-floral transition in many plants

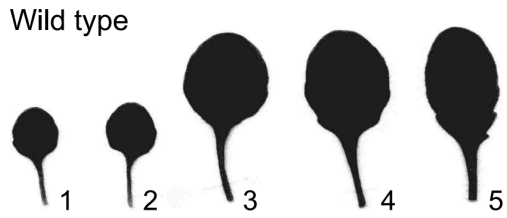


Figure 2.1: Phenotypic analysis of changes in leaf morphology in *A. thaliana* across the plant. Taken from Yu *et al.* 2013.

A juvenile-to-adult transition in *B. oleracea* has been documented in the literature (Salter and James, 1974; Hand and Atherton, 1987; Booij and Struik, 1990; Wurr *et al.*, 2016; Rosen *et al.*, 2018). However, the mechanism underlying this transition has not yet been elucidated. For Arabidopsis, there are some distinct morphological changes that distinguish the juvenile and adult phases of the plant life cycle. The juvenile leaves of Arabidopsis are rounded with smooth edges, however, adult leaves are more elongated with serrated edges (Fig. 2.1) (Willmann and Poethig, 2005; Bäurle and Dean, 2006; Wu *et al.*, 2009; Yu *et al.*, 2013). Furthermore, the adult leaves of Arabidopsis curl down at the edges and the blade-to-petiole ratio is higher than in juvenile leaves (Willmann and Poethig, 2005). Neither of these phenomena have been observed for *B. oleracea*.

In addition to these morphological markers, there are also key changes in gene expression, which mark the juvenile-to-adult transition in Arabidopsis. A decrease in *miRNA156/157*, leads to a corresponding increase in the expression of target genes encoding *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) transcription factors (Wu and Poethig, 2006; Wu *et al.*, 2009; Xu *et al.*, 2016; Guo *et al.*, 2017). The SPL proteins mediate both the morphological and physical transitions associated with the juvenile-to-adult transition in Arabidopsis, via the activation of *miRNA172* and the MADS-box genes (Fig. 2.2) (Wang, Czech and Weigel, 2009; Wu *et al.*, 2009; Yu *et al.*, 2013; Xu *et al.*, 2016). Although this work was carried out in Arabidopsis, both the *miR156* and *miR172* families have been identified in *B. oleracea* and *B. napus* (Buhtz *et al.*, 2008; Sunkar and Jagadeeswaran, 2008). Therefore, it is possible a decrease in *miR156* levels, could provide a genetic marker for the juvenile-to-adult transition in *B. oleracea.*

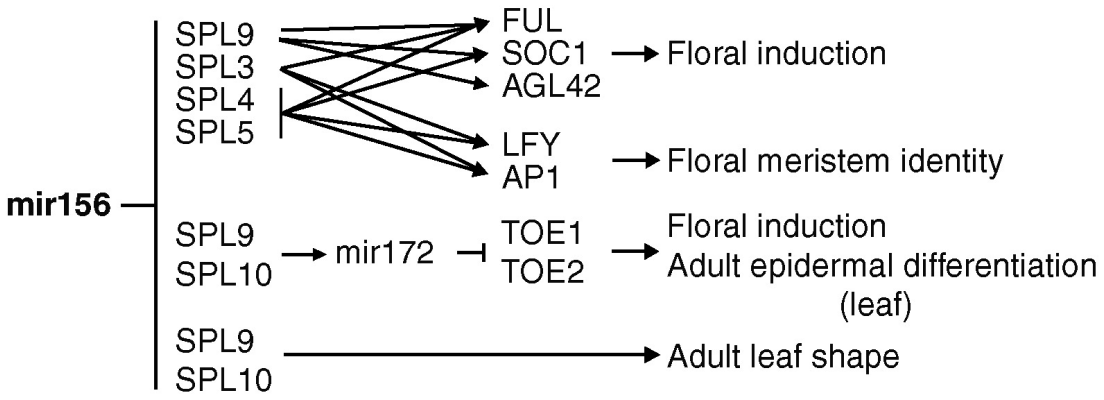


Figure 2.2: The downstream targets of the *miR156* and *miR172* families and the functions of those target genes, taken from Poethig, 2009.

The *miR156* family represses expression of the *SPL* gene family, whilst the *miR172* family represses expression of the *TOE* genes.

2.1.3 The vegetative-to-floral transition is critical to *B. oleracea* production

Unlike the juvenile-to-adult transition, the vegetative-to-floral transition has been studied extensively. For *B. oleracea,* the vegetative-to-floral transition is crucial for the production of quality vegetables. Various organs are harvested from *B. oleracea* subspecies, but for all the shift from the vegetative to reproductive phase is important and being able to genetically manipulate this transition will aid the development and production of synchronous brassica vegetables.

Despite this importance, much remains unknown about the floral transition within this *B. oleracea*. *B. oleracea* is a diploid, but it contains multiple copies of many genes due to a whole genome triplication event which occurred in an ancestral genome similar to that of Arabidopsis*,* around 23 Mya (Arias *et al.*, 2014). This genome triplication event gave rise to the modern Brassica species (Beilstein *et al.*, 2010).

Whilst the presence of multiple gene copies enables the emergence of novel gene functions by reducing both natural and artificial selective pressure on any one copy (Jones *et al.*, 2017), it also impedes efforts to translate knowledge of gene function from model to crop species (Conant and Wolfe, 2008). *B. oleracea,* contains orthologues of many Arabidopsis flowering time related genes, suggesting evolutionary conservation between the two species (Blümel,

Dally and Jung, 2015), which may aid the understanding of flowering time control in *B. oleracea* through the transfer of knowledge from Arabidopsis.

2.1.4 Associative transcriptomics is a robust method for identifying significant associations with target traits

Genome wide association studies (GWAS) have long been used as an effective means of identifying candidate genes for target traits. Linkage disequilibrium (LD) is the non-random association of alleles at different loci within a population. If the frequency of association between the different alleles of a locus is higher or lower than what would be expected if the loci were independent and randomly assorted, then the loci are said to be in LD (Flint-Garcia, Thornsberry and Buckler, 2003). LD will vary across the genome and across chromosomes and it is important to account for this in GWAS analysis. Selection, mutation rate and genetic drift, are some of the factors that lead to the variation in LD seen across genomes and strong selection or admixture within a population will increase LD. As a consequence it is important to account for population structure in GWAS analyses, which can be determined using unlinked markers (Evanno, Regnaut and Goudet, 2005).

GWAS allows for the identification of markers in LD with loci controlling traits. This has been an extremely important tool in human genetics and has been successfully used in numerous plant species, including Arabidopsis, maize, rice and Brassica (Huang *et al.*, 2012; Romero Navarro *et al.*, 2017; Havlickova *et al.*, 2018; Raman *et al.*, 2019). However, this is reliant on existing genomic resources for the species, which are often not available for many crops due to the complexity within their genome. Associative transcriptomics uses transcriptome sequence aligned to a reference, to identify and score molecular markers which represent variation in gene sequences and expression levels, and correlate this to trait data. Alignment of reads against a reference can be used to identify SNPs which are used as markers in GWAS. The number of reads aligned to a given gene provides the gene expression level, as RPKM (Reads per Kilobase of transcript per Million mapped reads) or FPKM (Fragments per Kilobase of transcript per Million mapped reads). Regression of the trait value against this expression level can then be used to determine gene expression markers (GEM). Associative transcriptomics is a robust method for identifying significant associations and is being increasingly used to identify molecular markers linked to trait controlling loci in crops (Rafalski, 2010; Harper *et al.*, 2012, 2016; Havlickova *et al.*, 2018).

2.1.5 Hypotheses and Aims

To investigate the effects of vernalisation timing, temperature, and duration on flowering time in a diverse panel of *B. oleracea,* a large-scale phenotyping experiment was designed and carried out. This chapter discusses the results from this experiment and goes on to use them to generate and validate a novel associative transcriptomics pipeline for *B. oleracea*. This pipeline is subsequently used to identify candidate genes for vernalisation response in *B. oleracea.*

With warmer winter conditions becoming more frequent here in the UK, it is important to identify lines that can still perform under these changing conditions in order to future proof our crops. Vernalisation is known to vary between accessions and is documented as being an adaptive trait, therefore I hypothesise that the comparison of the responses to different vernalisation treatments will identify *B. oleracea* lines which require less or no vernalisation for their development.

The shift from vegetative to floral growth is a key developmental transition in *B. oleracea*. Many of the genes involved in this transition have been identified in Arabidopsis, and orthologues of these genes are present in *B. oleracea*. Given the close evolutionary relationship between Arabidopsis and *B. oleracea*, I hypothesise that orthologues of many key Arabidopsis Flor-ID (Bouché *et al.*, 2015) genes will also be important in the vegetative-to-floral transition in *B. oleracea.*

Identifying candidate genes for traits associated with vernalisation response could provide key information for growers and breeders that could aid future breeding programs, information that could prove invaluable under current changing climatic conditions. Whilst orthologues of Arabidopsis Flor-ID genes are documented in *B. oleracea,* it should be noted that there are often multiple paralogues of these genes*,* due to an ancient genome triplication event in the evolutionary history of *B. oleracea*. *FLC* is known to have an important role in determining vernalisation requirement in both Arabidopsis and *B. oleracea* and much work has been carried out on one specific *B. oleracea FLC* copy, *BoFLC.C2* (Lin *et al.*, 2005; Irwin *et al.*, 2016). It is known that *cis* polymorphism at *BoFLC.C2* confers different vernalisation requirements within *B. oleracea* (Irwin *et al.*, 2016). I hypothesise *BoFLC.C2* will be identified as an important candidate for the phenotypic response to vernalisation using associative transcriptomic analysis, which will act to determine the efficacy of the pipeline.

2.2 Methods

2.2.1 Understanding the effect of vernalisation on the floral transition in *B. oleracea*

A subset of 69 lines fixed, as doubled haploids (DH) or at S4 and above, were chosen from the *Brassica oleracea* Diversity Fixed Foundation Set (Walley *et al.*, 2012) (Fig. S1A). This subset aimed to encompass as much variation by ecogeographic origin and crop type as possible. Seed was sown and seedlings pricked out into individual pots. Plants were watered daily, fed as required and given a pre-growth period of either six or ten weeks in a glasshouse. Natural light was supplemented with LED lights giving 16 h light (00:00 – 16:00) and 8 h dark. The daytime temperature was maintained at a minimum of 21 ºC day, 18 ºC night. At the end of the pre-growth period, three plants of each line for each of the pre-growth treatments were transferred to Conviron controlled environment rooms for a period of vernalisation at 5, 10 or 15 ºC, 16h LED light (00:00 – 16:00) and 8 h dark, 60% humidity. Vernalisation duration was either six or twelve weeks, after which plants were re-potted into 2 l pots and placed into a polytunnel using a randomised block design. Due to the staggered sowing for each experiment (Table 2.1), all plants came out of vernalisation and into the polytunnel on the same day. In the polytunnel, plants received natural light only. Additionally, three replicates of each line were grown without vernalisation as a non-vernalised control group. This resulted in a total of 1863 plants.

Table 2.1: Sowing and vernalisation schedule for phenotyping to investigate the effects of environmental perturbations on flowering time.

Sowing Date	Pre-Growth Length (wks)	Vernalisation Length (wks)	Into Vernalisation	Out of Vernalisation
28/12/2018	10	12	08/02/2018	03/05/2018
30/11/2018	6	12	08/02/2018	03/05/2018
11/01/2018	10	6	22/03/2018	03/05/2018
22/02/2018	10	-	-	03/05/2018

The leaf number of each plant was recorded the day before transfer to vernalisation. For experiments, 1, 3 and NV this was at 70 days after sowing and for experiment 2 this was on day 42. The plants were also scored for days to head, or buds visible (BBCH51) and upon opening of first flower (BBCH60), days to first flower. The BBCH Scale (Fig. 2.3) is a well-established scale for measuring development within both mono- and dicotyledonous plants (Meier, 2001) and was chosen for use to standardise scoring.

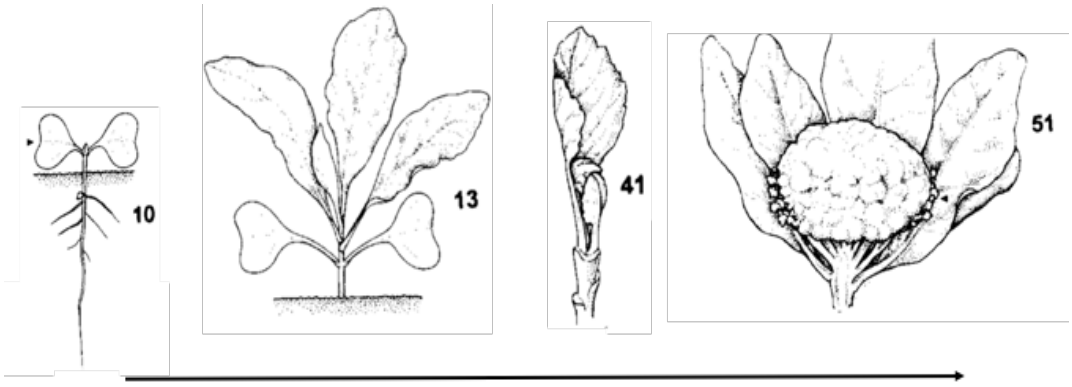


Figure 2.3: Key stages of the BBCH scale for Brassica vegetables.

10 cotyledons completely unfurled, 13 third true leaf fully expanded, 41 head begins to form, youngest two leaves remain folded to protect it. 51, buds fully visible when looking down on the plant. Adapted from Meier 2001.

Summary statistics were computed using R Version 3.6.3 (Team, 2013) as part of an initial analysis of the data. For comparison, the days in each treatment were removed from the final measures, leaving only days to BBCH51 and BBCH60 post vernalisation. Individuals which did not flower were given an arbitrary value which was the last scoring date, plus five days, post vernalisation, equating to 126 days. Some lines were removed from the analysis due to heterogeneity seen in their flowering times, these were GT120233, GT120234, GT120194 and GT120195. It is possible that these lines were not at a homozygous state and therefore have been removed from all analysis, resulting in 65 lines being taken forward. Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) (or relative entropy) was used to assess leaf numbers as a proxy for plant development, as temperature data was unavailable for the glasshouse used for the pre-growth period. KL uses a reference distribution to measure the difference with a second probability distribution, therefore it was well suited for comparing the distributions of leaf numbers between treatments. Here we were able to compare the distribution of leaf numbers across the population between experiments, to identify if there

were global differences in leaf number between treatments. Typically, plant age and leaf number (including the leaf scars of leaves which have fallen) are positively correlated, therefore the leaf number can be used as a proxy for plant age.

2.2.2 Associative transcriptomics to identify candidate genes controlling the phenotypic response to vernalisation

The SNP and expression data used for associative transcriptomics, was obtained from the group of Prof. Bancroft at the University of York. Leaf tissue was collected from 131 *B. oleracea* accessions and sent for RNA-Seq. A total of 110555 SNP markers were anchored to the C file of the *Brassica* pantranscriptome (He *et al.*, 2015) using Maq v0.7.1. SNP positions were excluded if they did not have a read depth in excess of 10, a base call quality above Q20, missing data below 0.25, and 3 alleles or fewer.

A population structure was generated from the 69 *B. oleracea* accessions (Fig. S.6A) used for phenotyping. Stringent rules were used to first prune the SNP data to ensure it only contained unlinked markers. SNPs were required to be biallelic, with a minor allele frequency (MAF) > 0.05, one per gene and a minimum distance of 500 bp apart. The Bayesian clustering algorithms implemented in the software STRUCTURE (Pritchard, Stephens and Donnelly, 2000) were then used to determine population structure with a burn-in of 10000, a MCMC of 10000 and 20 iterations. By analysing the differences in the distributions of genetic variants amongst populations, STRUCTURE is able to identify groups with similar patterns of genetic variation. It works by applying a model to the data with *K,* an assumed number of groups that are defined by a subset of allele frequencies obtained from the population. The model cycles through a defined number of *K* values, and the likelihood of each *K* value being a good descriptor of the data is calculated using posterior probabilities of *K.* A Q-matrix can then be calculated. Each individual is assigned membership coefficients, which sum to one for each group; the individual is assigned to the group for which it has the highest probability. The Q-matrix is the average individual membership coefficients to each group, calculated from all iterations of the *K* value chosen. STRUCTURE HARVESTER (Earl and vonHoldt, 2012) was used to determine the optimal *K* value for the population and to produce the Q matrix (Table S.4A) used in GWAS analysis.

TASSEL (Bradbury *et al.*, 2007) version 5.0 was used to conduct the most appropriate model for each trait based on the QQ plots generated (Fig. S.8A). Generalised linear models (GLM), both with and without correction for population structure using the Q matrix and PCA, and

mixed linear models (MLM) using kinship data calculated using TASSEL's 'centered IBS' method was utilised to look for associations. Optimum compression level and P3D variance component estimation were used as MLM options. For GWAS analysis only SNP markers with an allele frequency > 0.05 were used. As a positive control mean percentage oil content of seeds were used, following validation of the pipeline phenotypic data collected as part of the vernalisation experiment was used to look for associations. Pairwise linkage disequilibrium was calculated for each chromosome using the SlidingWindow function within TASSEL version 5.0. R^2 values was plotted against the base pair distance (Fig. S.9A). SNPs were removed from GWAS analysis if their MAF was below 0.05. TASSEL was used to construct phylogenetic trees, using the Neighbour Joining method (Saitou and Nei, 1987) and all SNPs with MAF > 0.05. Trees were graphed in R using the package ggtree (Yu *et al.*, 2017).

Gene Expression Marker (GEM) associations were calculated by an in-house script in R Version 3.6.3 using a fixed effect linear model with RPKM values. The script removed all markers with an average expression below 0.5 RPKM and performed a linear regression using RPKM as a predictor value to predict a quantitative outcome of the trait value. Regression is a statistical method used to determine the relationship between a dependent variable and a one or more independent variables, in this case determining the relationship between a locus and a series of traits. The aim is to build a mathematical model to define the trait as a function of RPKM and if this model is statistically significant, determined by ANOVA, both SNP and GEM outputs were plotted as Manhattan Plots created using an in-house R script.

Significance for both GWAS and GEM analyses was determined using the false discovery rate (FDR). FDR is calculated on a trait-by-trait basis and uses the P-values from the GWAS or GEM analysis to determine the proportion of false discoveries across all discoveries in the experiment (Benjamini *et al.*, 2005). FDR was calculated using the QValue package in R (Storey *et al.*, 2019)

2.2.3 DNA Extraction

Genomic DNA of accessions used in EcoTILLING was prepared from young leaf tissue of plants grown in a glasshouse. Plants were watered daily, fed as required and natural light was supplemented with LED lights giving 16 h light (00:00 – 16:00) and 8 h dark. The daytime temperature was maintained at a minimum of 21 °C day and 18 °C at night. Light was excluded for 48 h prior to harvesting. Approximately 3 g of tissue was used per accession and nuclei extraction buffer (10 mM Tris-HCL pH 9.5, 10 mM EDTA pH 8.0, 100 mM KCL, 500 mM

sucrose, 4 mM spermidine, 1 mM spermine, 0.1 % 2-mercaptoethanol) was added. The solution was vortexed and filtered through two layers of Miracloth. Lysis buffer (10 mM Tris-HCL pH 9.5, 10mM EDTA pH 8.0, 100 mM KCL, 500 mM sucrose, 4 mM spermidine, 1 mM spermine, 0.1 % 2-mercaptoethanol, 10 % Triton-x) was added to the homogenate, before centrifuging at 1000 g for 20 min at 4 °C. Supernatant was removed and the pellet resuspended in CTAB extraction buffer (100 mM Tris-HCL pH 7.5, 0.7 M NaCl, 10 mM EDTA pH 8.0, 1 % cetyl trimethylammonium bromide, 1 % 2-mercaptoethanol). The solution was incubated at 60 °C for 30 min. An equal volume of chloroform:isoamyl alcohol (24:1) was added and the solution rotated for 10 min before centrifugation for 10 min at 1000 g. The aqueous phase was removed and RNase T1 (50 units/ml) and RNaseA (50 µg/ml) added. The solution was incubated at 37 °C for 45 min. Proteinase K (150 µ/ml) was added and the solution incubated at 37 °C for a further 45 min. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and the solution centrifuged for 10 min at 1000 g, this was repeated once more. An equal volume of chloroform:isoamyl alcohol was added and the solution rotated for 10 mins, before centrifuging at 1000 g for 10 min. DNA was then precipitated by the addition of sodium acetate (10 % 3 M) and 3 x volume of 100 % ethanol. The solution was centrifuged at 1000 g for 1 min to gently pellet the DNA and it was subsequently washed with 3 ml of 75 % ethanol. The solution was gently centrifuged, the ethanol poured off and the pellet left to air dry, before being resuspended in 50 µl dH_2O. DNA was checked for quality. Nanodrop analysis was carried out on 1.5 µl of each DNA sample and the extracted DNA was stored at -20 °C until required.

2.2.4 Targeted Sequence Enrichment analysis

A bait library for targeted sequence enrichment for a specific subset of genes was developed and synthesized with Arbor Biosciences (https://arborbiosci.com/). DNA was extracted and sent to Arbor for enrichment and Illumina sequencing. Reads from individual accessions were mapped to the reference sequence of *B. napus* cv. Darmor-*bzh* (Chalhoub *et al.*, 2014) using BWA (Li and Durbin, 2009) version 0.7.17-r1188. For BWA, we used aln/sampe and standard parameters. Mapped reads were sorted and indexed using SAMTOOLS (Li *et al.*, 2009) version 1.10, sort and index and subsequently visualized with Integrative Genomics Viewer (IGV)(Robinson *et al.*, 2011).

2.3 Results

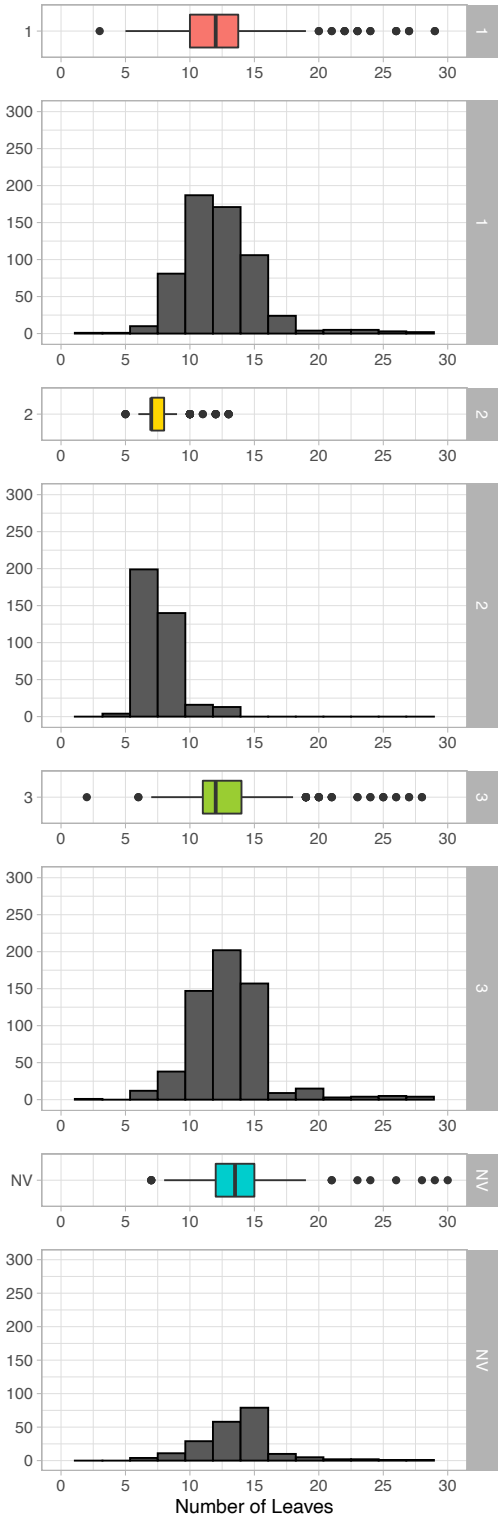


Figure 2.4: Boxplots with corresponding histograms to demonstrate the distribution in leaf numbers between experiments prior to vernalisation.

For boxplots, central box indicates interquartile range, black dots extremities..

2.3.1 Leaf number correlates with plant age amongst the phenotyped accessions

Four different treatments were designed to assess the effects of the timing of vernalisation, vernalisation length and vernalisation temperature on the vegetative-to-floral transition in *B. oleracea*. Temperature data for the glasshouse used for the pre-growth period was unavailable, therefore the leaf data collected prior to vernalisation was used as a proxy for plant development. This was important because sowing was staggered and therefore the plants were growing in the glasshouse at different times, which may have resulted in exposure to different temperatures that could have led to more favourable growth conditions for some sowings. The leaf numbers recorded for experiments 1, 3 and NV were all taken at 70 d, however, the leaf numbers taken for experiment 2 were collected at 42 d, as this was the experiment with the shorter pre-growth period. Consequently, it was predicted that leaf numbers for all experiments should be similar despite the staggered sowing dates, excluding for experiment 2, in which it was predicted there would be fewer leaves due to the younger age of the plants. Indeed, this was seen to be the case (Fig. 2.4), the mean leaf numbers for experiments, 1, 3 and NV were 12.2, 12.8 and 13.7 respectively, whereas for experiment 2 this was lower, at a mean of 7.6 leaves. The distribution of leaf numbers looks very similar for most experiments, the Kullback-Leibler (KL) divergence was used to compare the difference between the distributions in leaf numbers.

The KL divergence between experiment 1 and NV was 0.268 bits, between experiment 3 and NV as 0.135 bits, however the comparison between experiment 2 and NV was 3.324 bits, reflecting the differences seen in mean leaf numbers and supporting the conclusion that experiments 1, 3 and NV are comparable, as is experiment 2 as there is a clear developmental difference with this set of plants in comparison to the others.

2.3.2 Perturbing the timing of vernalisation enabled the identification of lines with a juvenile phase

B. oleracea is reported to have a juvenile phase in the literature (Juvenility, Hand and Atherton, 1987; Booij and Struik, 1990; Walley *et al.*, 2012) and has also been documented by growers and breeders not to flower if placed into the cold too early (Tallis, 2011), indicating they must need reach a certain age or developmental state before being able to respond to floral inductive cues and transition to the floral phase. Consequently, it was expected that the longer pre-growth would result in a larger number of flowering individuals, as an increased number of individuals should be able to respond to cold as a floral inductive cue, due to the increased developmental stage of the plants. There was no significant difference in the numbers of individuals reaching BBCH51 after the ten-week pre-growth compared to the six-week pre-growth, at 179 and 167 individuals respectively (Two-proportion Z-Test, $\chi^2 = 3.534$, df = 1, $P = 0.060$). Despite the lack of difference in the counts, the mean days to BBCH51, post vernalisation, was significantly higher after exposure to the six-week pre-growth period at 21.0 d (Wilcoxon Test, W = 17958, $P = 0.004$), compared to 5.8 d following the ten-week pre-growth period. This significant difference in mean days to head between the pre-growth periods, suggest that some individuals may be demonstrating a juvenile phase, and therefore be unable to respond to vernalisation as a floral inductive cue following the shorter six-week pre-growth period, resulting in no flowering or later flowering compared to those individuals exposed to a ten-week pre-growth period.

This was further reflected in comparisons of the mean days to BBCH60. The mean days to BBCH60 was 57.9 d after the six-week pre-growth period and was significantly lower, at 35.9 d following the ten-week pre-growth period (Wilcoxon Test, W = 17471, P = 2.96e-05). Despite the differences seen in the mean days to BBCH51, the distribution of heading over time between the two treatments was very similar (Fig. 2.5), with heading mainly focused around the first few weeks after return to warm conditions.

Plotting the results as a heatmap allowed exploration of the entire dataset on a line-by-line basis (Fig. 2.6). For 47 lines, the mean days to BBCH51 was earlier after the ten-week pre-growth period than the six-week period, indicating these lines were potential candidates for juvenility. Of these 47 lines, only seven were significantly earlier ($P < 0.05$), with all seven not flowering following the six-week pre-growth period (Table 2.2).

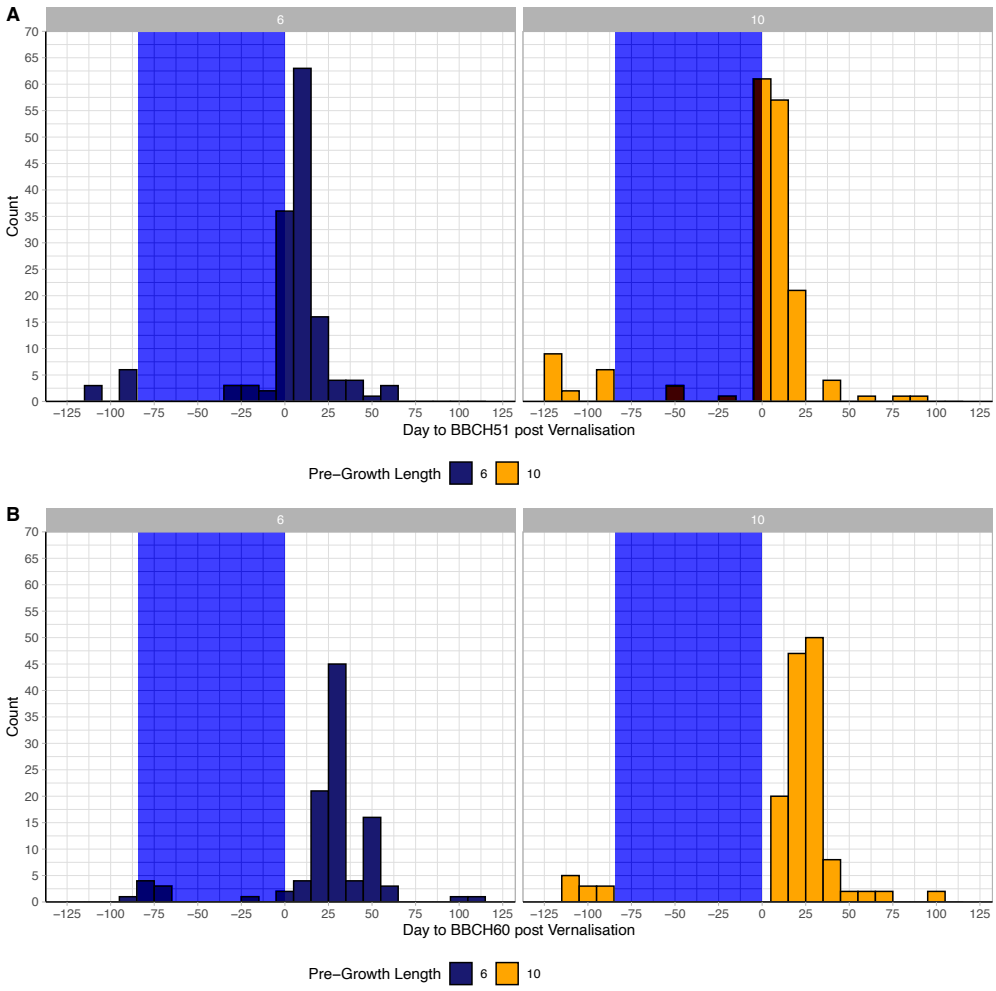


Figure 2.5: No significant difference was seen in the distribution of days to BBCH51 or days to BBCH60 after exposure to a six or ten-week pre-growth period.

Heading dates post vernalisation, anything with a value below zero reached BBCH51/BBCH60 during or before vernalisation. Individuals that did not reach BBCH51 have been removed from this plot. Blue area represents time spent in vernalisation.

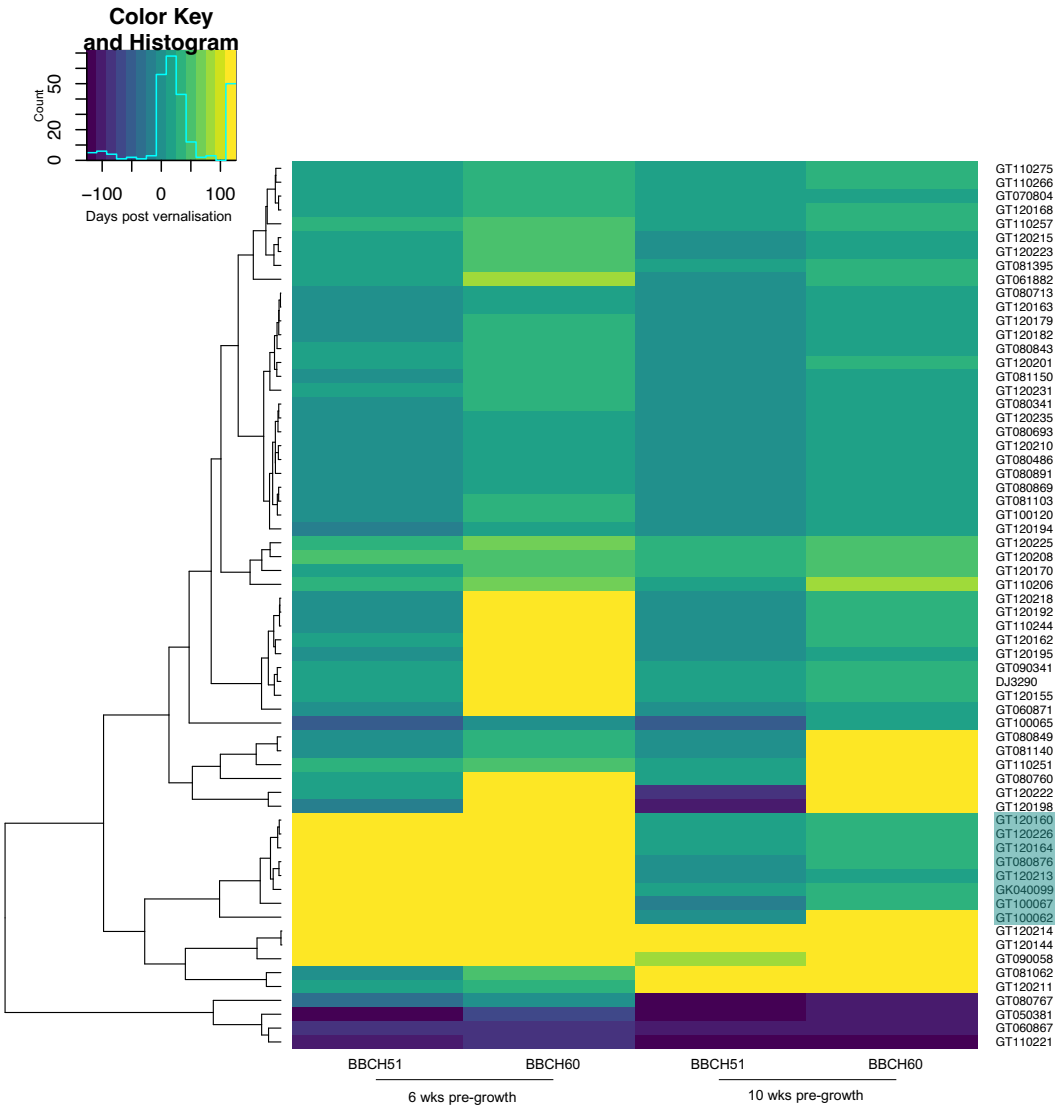


Figure 2.6: A longer pre-growth period led to earlier heading in multiple lines. Heatmap, blue indicates lines that went to head after vernalisation, and yellow before exposure to vernalisation at 5°C. Hierarchical clustering results based on flowering phenotype shown on the left of the heatmap. Highlighted lines are candidates for juvenility.

Table 2.2: Seven lines went to head significantly earlier following the ten-week pre-growth period.

		Mean days to BBCH51	
Genotype	**Crop Type**	**6 wk pre-growth**	**12 wk pre-growth**
GK040099	Cauliflower	126	9
GT080876	Cauliflower	126	6.33
GT100062	Broccoli	126	2
GT100067	Kale	126	-12
GT120160	Calabrese	126	17
GT120164	Broccoli	126	22.33
GT120213	Cauliflower	126	8
GT120226	Cabbage	126	18

2.3.3 Vernalisation length is a key determinant of the vegetative to floral transition

In addition to the timing of vernalisation, the effect of vernalisation length was also analysed. Minimal difference was seen between the numbers of individuals that reached BBCH51 after the two vernalisation lengths, with just ten individuals not reaching BBCH51 after the six-week vernalisation treatment and thirteen individuals after the twelve-week treatment (Two-proportion Z-Test, $\chi^2 = 0.185$, df = 1, $P = 0.667$). Despite this, a significant difference was seen in the mean days to BBCH51 after exposure to the two treatments. After six-weeks of vernalisation, the mean days to BBCH51 was 9.5, this was significantly reduced after exposure to twelve-weeks to 5.8 d (Wilcoxon Test, W = 19532, $P = 0.002$). This reduction in mean days to BBCH51 was coupled with more synchronous heading between lines (Fig. 2.7). This was an expected outcome, as with a longer vernalisation period we would predict that the vernalisation requirement of more lines would be met, and as a consequence they would be able to transition from a vegetative to a floral state during the course of the experiment. However, when comparing the mean days to BBCH60, no significant difference was found between the two treatments (Wilcoxon Test, W = 18195, $P = 0.006$). The mean days to BBCH60 following the six-week vernalisation treatment was 40.4 d and after the ten-week treatment was 35.9 d. Indicating although the transition was faster, flowering itself was slowed, perhaps to prolong the period over which the plant was flowering. Alternatively, this could be in response to the post-vernalisation environment. Temperatures reached in excess of 40 ºC in the polytunnel on the plants return to warm (Fig. S.3A and S.4A).

Distinct phenotypic groupings are evident when looking across the population (Fig. 2.8). The rapid cycling varieties are all clearly grouped at the bottom of the heatmap, reaching BBCH51 prior to exposure to vernalisation. For the majority of accessions, the difference in days to BBCH51 between the two treatments is apparent, but minimal. A more pronounced effect would be expected if vernalisation length alone was the key player in affecting the heading of *B. oleracea.* Despite this, two accessions, both cabbages, were the only lines which did not flower after the six-week vernalisation period but did after the twelve-week vernalisation. This phenotype indicates that these lines have an obligate vernalisation requirement and interestingly they were the two lines which were identified as good candidates for juvenility. These lines were GT090058 and GT120226.

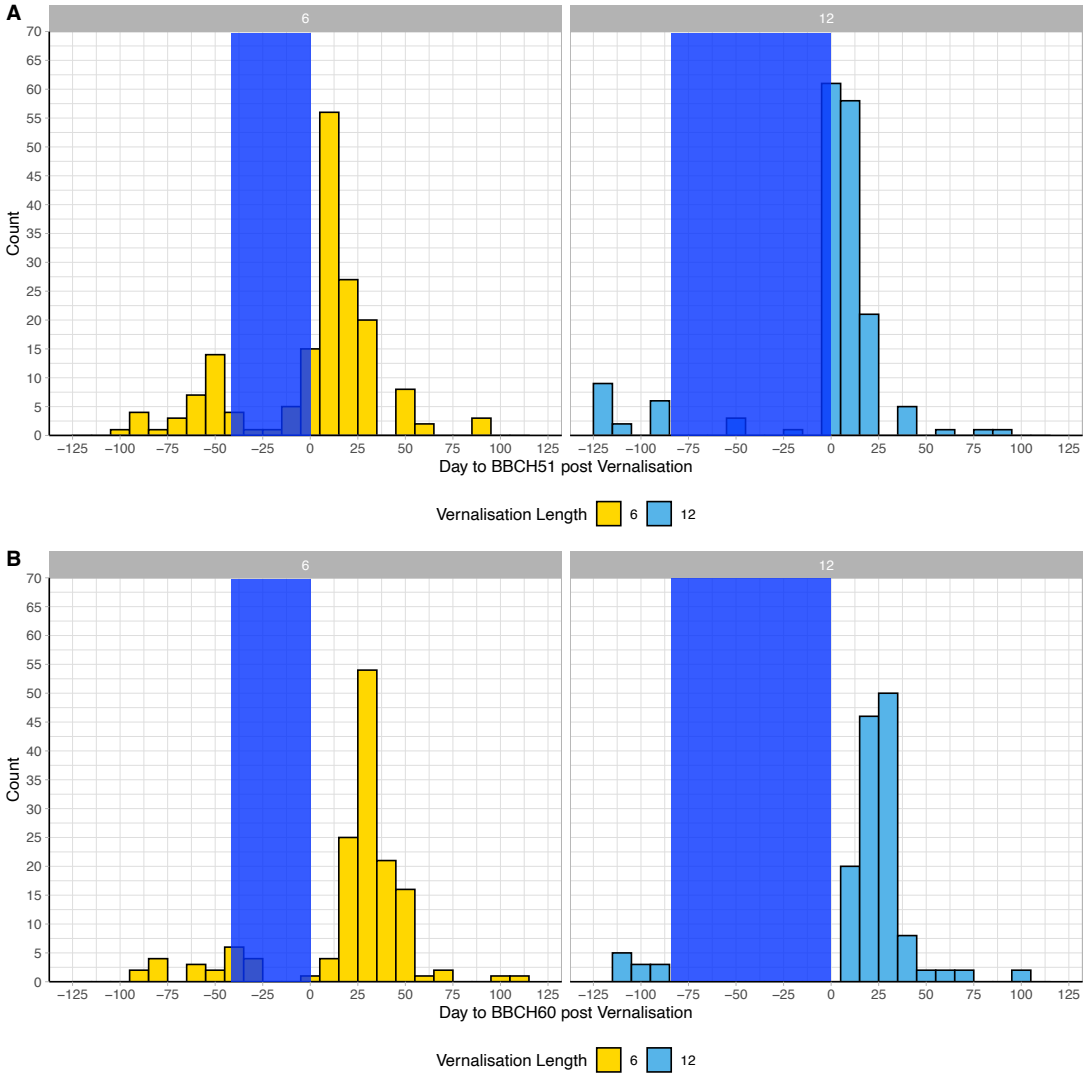


Figure 2.7: Exposure to twelve-weeks of vernalisation resulted in more synchronous heading across the population.

Histograms representing the distribution of heading and flowering dates across the population after exposure to either a six or a twelve-week pre-growth period. Heading dates post vernalisation, anything with a value below zero reached BBCH51 before or during vernalisation. Individuals that did not reach BBCH51 have been removed from this plot. Blue area represents time spent in vernalisation.

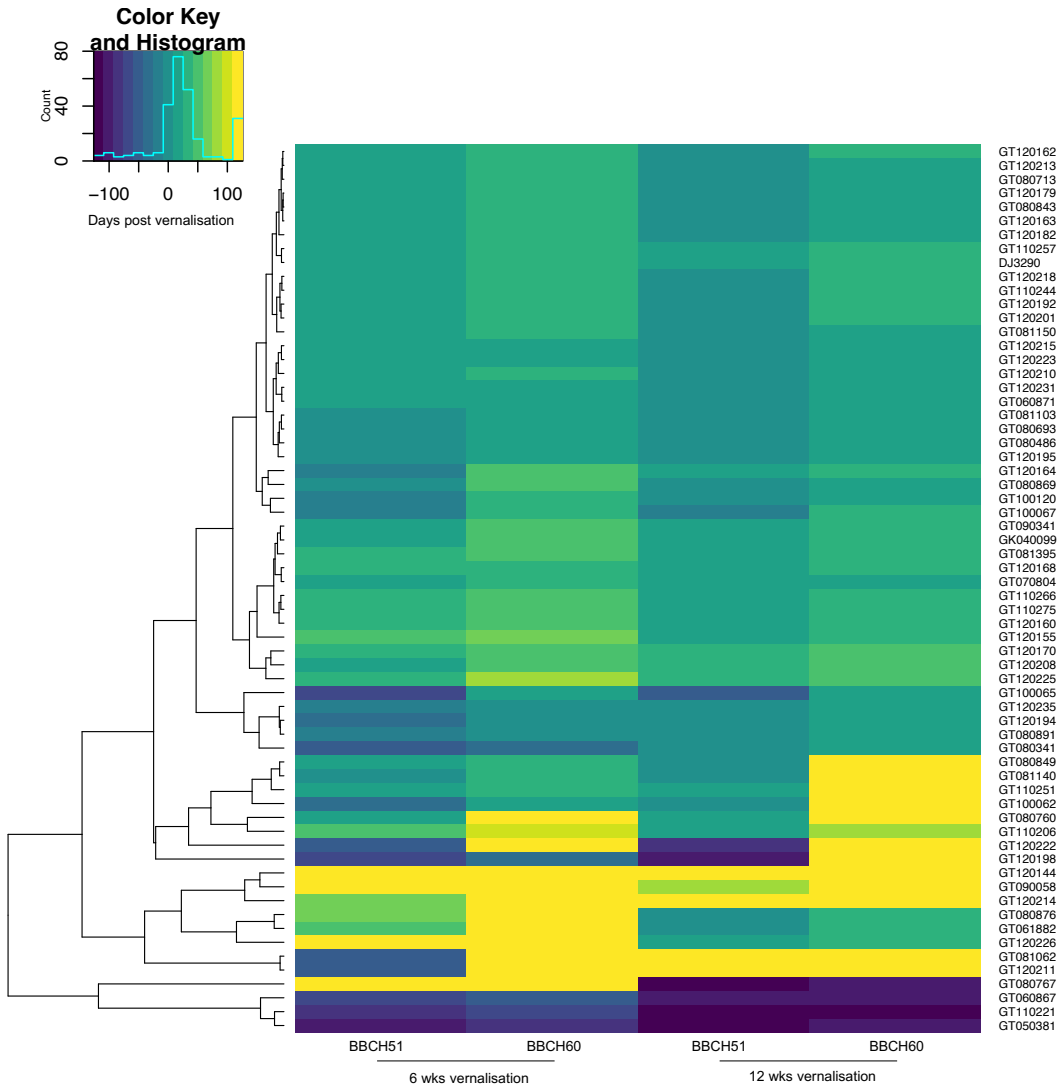


Figure 2.8: For the majority of lines, vernalisation length had a minimal effect on days to head and days to flower.

Heatmap, blue indicates lines that reached BBCH51 after vernalisation, and yellow before exposure to vernalisation at 5 ˚C. Variation in response to vernalisation length is genotype dependent, with most of the genotypes tested reaching buds visible faster after the longer vernalisation treatment. Hierarchical clustering results shown on the left of the heatmap.

2.3.4 Altering vernalisation temperature results in varying phenotypic response between accessions

Vernalisation temperature was the final environmental perturbation to be assessed. The warmest vernalisation temperature, 15 ºC, had the highest instance of plants not reaching BBCH51, at 24 individuals: excluding those which died or contracted disease during the

course of the experiment. Vernalisation at 5 and 10 ºC resulted in 13 and 12 individuals not reaching BBCH51 respectively. There was found to be no significant difference in the numbers that reached BBCH51 across all three temperatures (Chi-Squared Test, $\chi^2 = 5.933$, df = 2, $P = 0.052$). Differences were seen in the mean days to BBCH51 between temperature treatments. The coolest vernalisation temperature, 5 ºC, gave the highest mean days to BBCH51 at 5.8 d, in comparison to -7.7 d after 10 ºC and 2.2 d after 15 ºC. No significant difference was seen in mean days to BBCH51 between 10 and 15 ºC (Wilcoxon Rank Sum Test, P=0.317), but the mean days to BBCH51 after a 5 ºC vernalisation treatment was significantly higher than both 10 and 15 ºC (Wilcoxon Rank Sum Test, W = 2.4e-06, $P = 0.009$). This indicates that the coolest vernalisation temperature had the largest overall effect on the flowering phenotypes. For days to BBCH60, the 10 ºC treatment appeared to have the largest effect on phenotype, with the mean days to BBCH60 being significantly higher after this treatment, at 43.2 d, compared to 35.9 d after 5 ºC and 31.2 d after 15 ºC (Wilcoxon Rank Sum Test, W = 0.001, $P = 0.047$). Therefore, in this study it took longer for bud opening to begin following warmer vernalisation temperatures.

Furthermore, there were clear differences in the distribution of heading dates between the three temperature treatments. As vernalisation temperature was increased, the variation in days to BBCH51 and BBCH60 also increased, with 5 ºC giving the most synchronous heading and flowering across all lines (Fig. 2.9). Warmer temperatures increased the variation in heading and flowering dates, indicating that the cooler vernalisation temperature aided faster transitioning of some lines, but delayed the development of others. This was to be expected, as it is known that temperature is one of the key external variables that influence the timing of the vegetative-to-floral transition (Putterill, Laurie and Macknight, 2004; Rosen *et al.*, 2018).

The effect of vernalisation temperature on the floral transition is demonstrated clearly between the broccoli GT110244 and the brussels sprout GT120168 (Fig. 2.10). For GT110244, as the vernalisation temperature increased so did the rate of development. The plant exposed to vernalisation at 5 ºC exhibits a clear broccoli phenotype, with the distinctive compact head, however both the 10 and 15 ºC plants have bolted and the phenotype we typically associate with this crop type is lost. In contrast, for GT120168 temperature has had the opposite effect. The plant which received the warmest vernalisation temperature is the least developed, there is no elongation, and it is clearly still in a vegetative state. This individual did not flower before the end of the experiment; therefore, it can be concluded that this line has a strong vernalisation requirement which could only be saturated by cooler vernalisation temperatures.

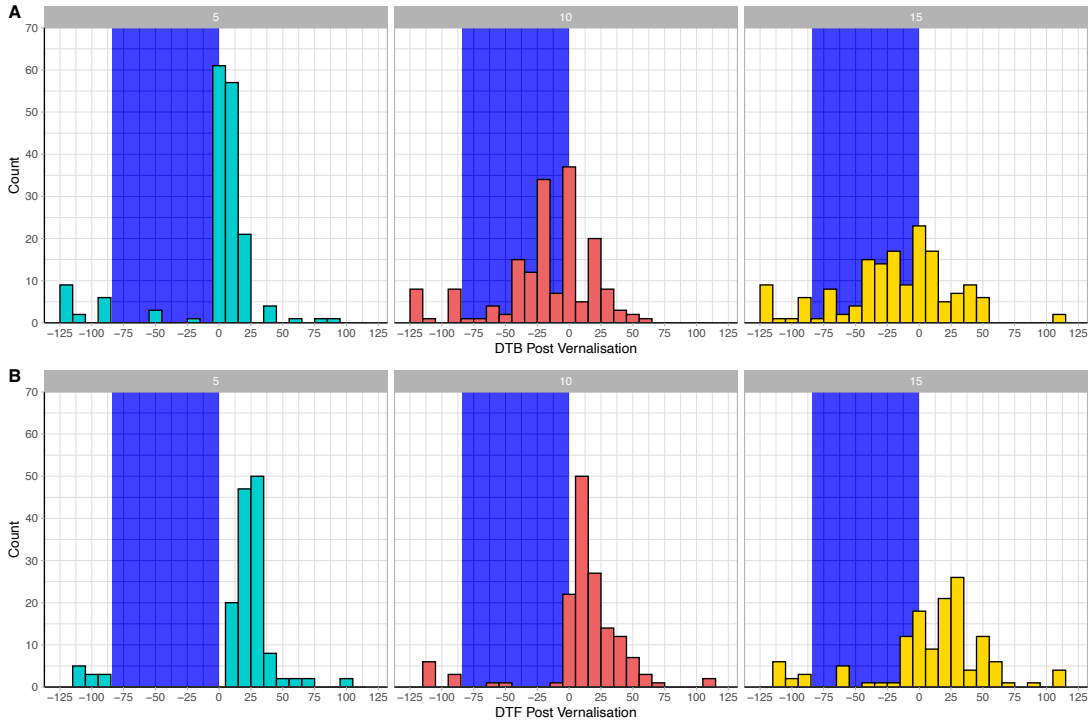


Figure 2.9: Increased synchrony in days to BBCH51 and days to BBCH60 was observed as vernalisation temperature was reduced.

Histograms representing the distribution of days to BBCH51 and BBCH60 post-vernalisation across the population after exposure to vernalisation at 5, 10 or 15 ºC. Individuals that did not reach BBCH51 have been removed from this plot. Blue area represents time spent in vernalisation.

There were multiple lines which were identified as having a strong vernalisation requirement like GT120168. Six more accessions did not reach BBCH51 after the warmest vernalisation treatment, indicating that 15 ºC for 12 weeks was not sufficient to saturate their vernalisation requirement and induce flowering. This group included three cauliflowers, one kale, one brussels sprout and one cabbage. The cabbage again was GT120226, reinforcing that this accession is a late flowering, strong vernalisation requiring line.

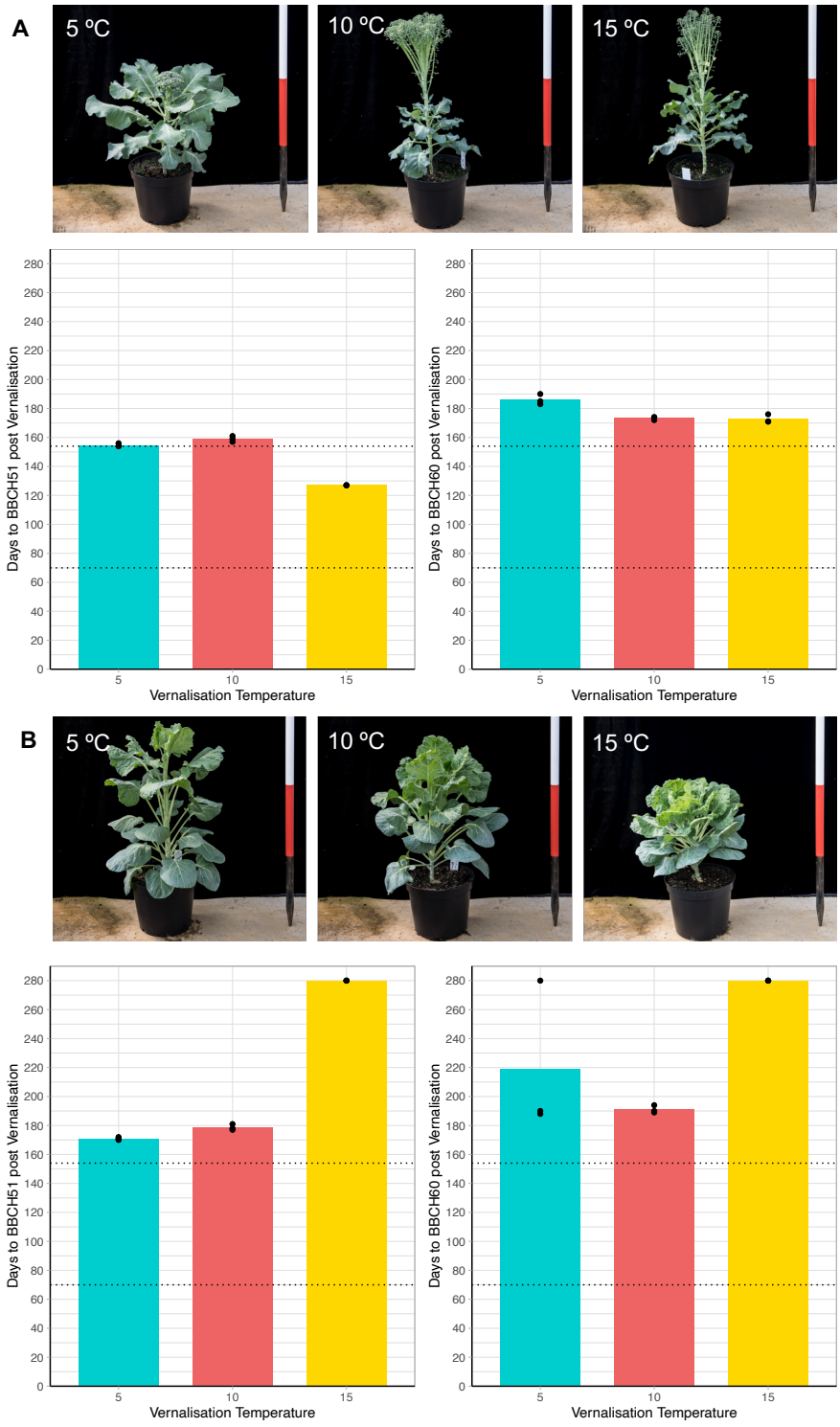


Figure 2.10: The phenotypic effect of vernalisation temperature is genotype dependent. A) GT110244, Broccoli, has developed at a faster rate after exposure to the warmer vernalisation temperatures. B) In contrast GT120168, Brussels Sprout, has developed at a quicker rate after exposure to the coldest vernalisation temperature assessed, 5 °C. The coloured bar represents the mean heading/flowering date for the three replicates at each treatment and the black dots represent individual heading/flowering dates.

2.3.5 Exposure to no vernalisation led to increased time to reach BBCH51 and BBCH60

A set of individuals were grown without exposure to vernalisation as a control, and to identify individuals that are not reliant on vernalisation for their transition from the vegetative to floral state. Of the 65 lines examined, only, seven did not reach BBCH51 or BBCH60 during the experiment. Unsurprisingly, due to the late flowering nature, both brussels sprout lines fell into this group, alongside a late flowering kale, a winter cauliflower and three of the broccoli lines. However, both GT120226 and GT090058, identified as candidates for an obligate vernalisation requirement, due to their response to the vernalisation temperature and length treatments were not in this group. They were, however, amongst the eleven lines which had at least one individual reach BBCH51 under non-vernalising conditions during the experiment, but not BBCH60. This reinforces the idea that these lines have a strong vernalisation requirement, but maybe this is a facultative requirement.

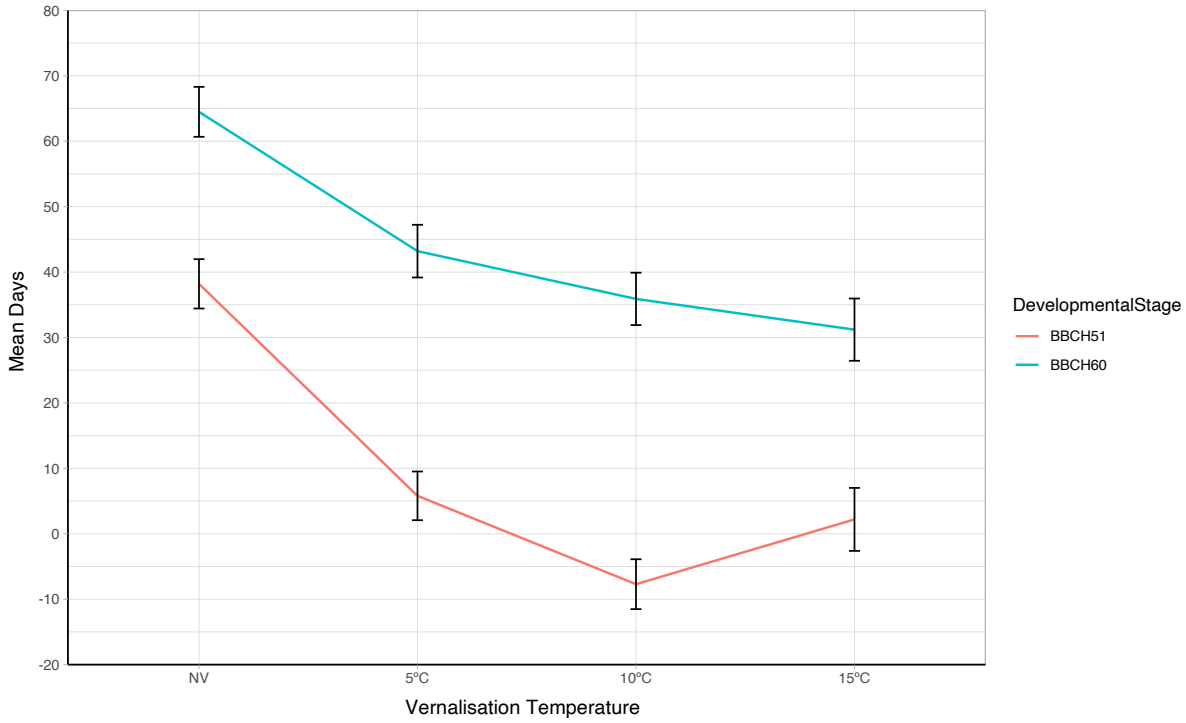


Figure 2.11: Vernalisation reduces the mean days to BBCH51 and days to BBCH60 across the population.

Pink line represents mean days to BBCH51, Blue represents days to BBCH60. The bars represent standard error.

Comparison of the mean days to BBCH51 and BBCH60 following vernalisation at 5, 10 or 15 ºC for twelve weeks to those for the non-vernalised set of individuals, demonstrated that with vernalisation comes on average a faster transition to the floral phase for this population (Fig. 2.11).

2.3.6 An associative transcriptomics pipeline was developed and validated for *B. oleracea*

Following initial statistical analysis to determine the phenotypic effects of the environmental perturbations on the vegetative-to-floral transition, associative transcriptomics was conducted to identify candidate genes which could account for these observations. Initially, associative transcriptomic analysis was carried out by the University of York (L. Havlickova) but did not yield any viable candidates.

GWAS requires trait data, SNP data and population data. The correct population data is extremely important, as an incorrect population structure may mean that significant associations may be identified on account of relatedness within the population, as opposed to association with a trait of interest. Many of the QQ plots displayed over correction and breaks, indicating that the population structure may not be appropriate for this dataset. As a result, the population structure was investigated.

The population structure used for the original analysis was produced using SNP data for 131 *B. oleracea* accessions, including the 69 used in this study, however when examining the clustering of those 69 lines within the population structure, it was clear the accessions were not grouped with those with similar phenotypes (Fig. 2.12E). Grouping by phenotype would imply that the groups are also genetically similar and of the same crop types. Calculation of ΔK showed a maximum value of $K = 2$, although a further peak was seen at $K = 5$ (Fig. S.7A) indicating there was substructure within the population. $K = 2$ is frequently seen as the top level of hierarchical structure, even when more sub-populations are present [21, 22]. Subsequent phylogenetic analysis (Fig. 2.13) identified clusters representing these subpopulations. Consequently, it was decided that the population structure required reassessment and would be produced specifically for the 69 accessions used in the phenotyping experiment only.

In addition to being generated using more accessions, the SNPs used to generate the original population structure required only a MAF score > 0.05 resulting in 36,631 SNPs, therefore, to generate a new population structure, more stringent requirements were used to select SNPs to ensure they were unlinked. STRUCTURE assumes markers are at LD with one another and are unlinked (Pritchard, Stephens and Donnelly, 2000). SNP calling in polyploids is notoriously difficult and the need to distinguish homoeologous SNPs from allelic SNPs teamed with the genetic similarity between sub genomes slows the rate of SNP marker development (Trick *et al.*, 2009; Clevenger *et al.*, 2015). Consequently, SNPs were required to be biallelic, as biallelic markers are the most robust and less likely to suffer from scoring errors. They also needed a MAF > 0.05 and a minimum distance of 500 bp between markers. This filtering process resulted in 664 SNPs which were used to generate a population structure, giving an average distance between markers of ~ 700 bp. Calculation of ΔK showed a maximum value of $K = 4$ (Fig. S.7A).

We assessed the two population structures based on crop type and phenotypic data. Using $K = 5$, generated using less stringent parameters, (Figs. 2.12A, 2.12C, 2.12E) cluster one contained only broccoli and calabrese, both members of the same subspecies var. *italica* (Labana and Gupta, 1993; Maggioni *et al.*, 2010), whilst cluster two was mainly cauliflower, subspecies var. *botrytis*. Some late flowering accessions were also included in both cluster one and two. This population structure grouped the rapid cycling and late flowering kales together with a spread of accessions from other crop types, into cluster four. The remaining two clusters were small by comparison: cluster three contained seven accessions, a mixture of broccoli, cauliflower, and kale, whereas cluster five consisted of just two lines, one kale and one cauliflower.

Using $K = 4$, generated using more stringent SNP selection criteria (Figs. 2.12B, 2.12D, 2.12F), cluster one contained all the rapid-cycling kales, characterised by their early heading and flowering phenotypes. Cluster two was mainly broccoli and calabrese, and cluster three consisted largely of the earlier flowering cauliflowers. Cluster four contained late flowering individuals across all seven crop types within the population, hence the larger variation in heading and flowering for this cluster. These groupings appeared to be more logical based on crop type and phenotypic information on the lines and phylogenetic analysis of the population further validated these findings (Fig. 2.13) giving tighter clustering, with clear groupings and only few outliers which demonstrated admixture, explaining their positions. Consequently, this population structure was taken forward and used in subsequent GWAS analysis.

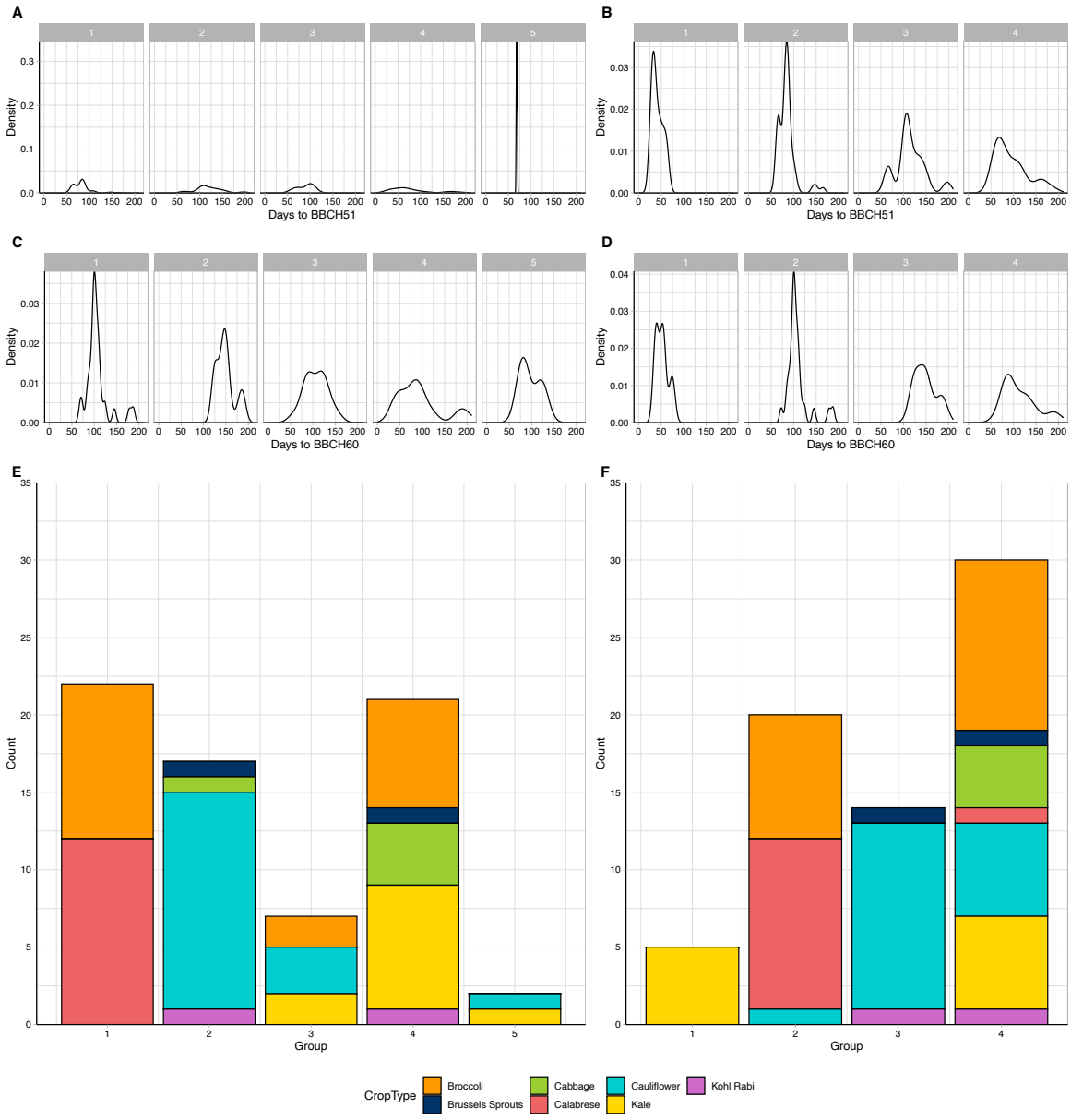


Figure 2.12: The rules for marker selection can significantly change the inferred population structure.

Density plots representing A) BBCH51, C) BBCH60 for the accessions within the five subpopulation clusters. Density plots representing B) BBCH51, D) BBCH60 for the accessions within the four subpopulation clusters. E) Population structure generated from SNPs with MAF > 0.05 F) Population structure generated from more stringent SNP pruning (Biallelic only, MAF > 0.05, > 500 bp apart, one per gene).

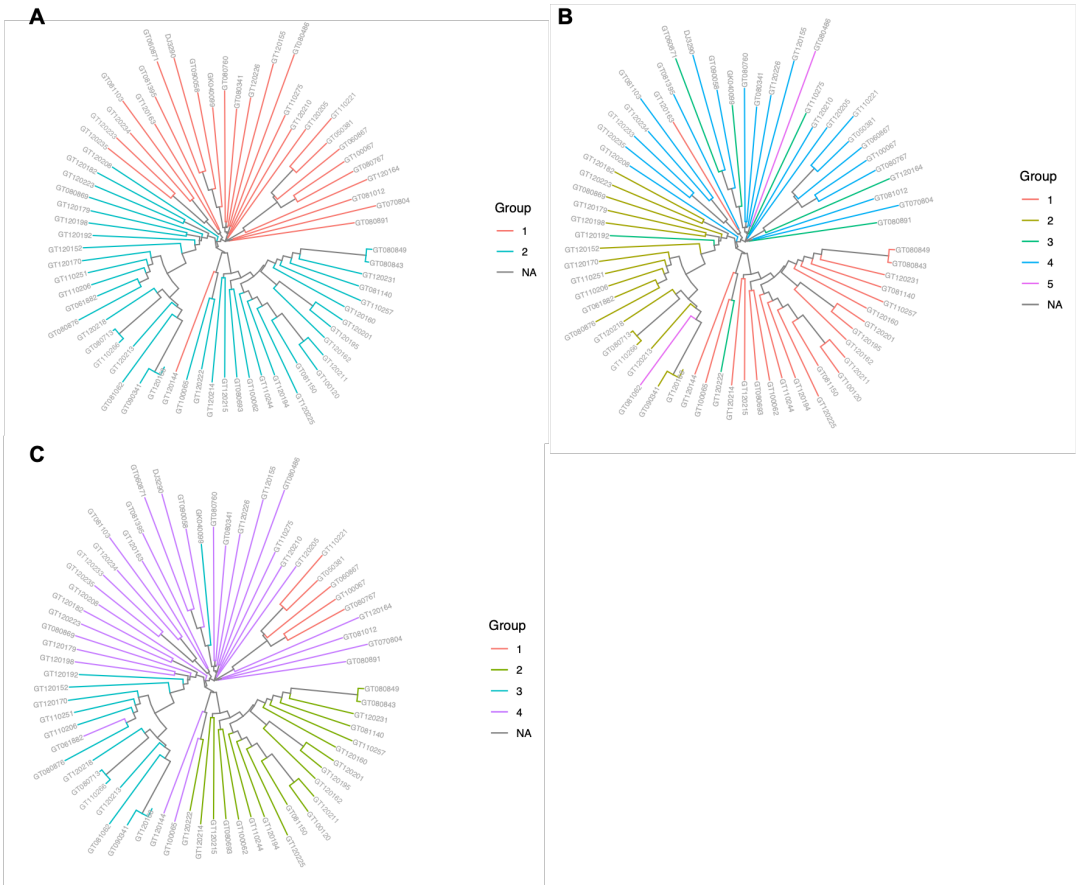


Figure 2.13: Association trees to demonstrate the substructure present within the phenotyped population.

A) $K = 2$, the highest level of structure seen within the population following analysis with the relaxed SNP set, B) $K = 5$, the substructure present within the population following analysis with the relaxed SNP set. C) $K = 4$, the result following population structure analysis on the stringent SNP set. Trees, generated in TASSEL using the Neighbour Joining method.

To further investigate the more stringent SNP selection criteria, a population structure was generated for all 131 accessions using these SNP requirements. This resulted in just 100 SNPs from which to construct a population structure, a low number. Analysis revealed a $K = 4$, matching that of the population structure generated for the 69 phenotyped accessions. The groupings of the 69 lines within these 131 were very similar, further adding confidence to the population structure.

To gauge the extent of linkage disequilibrium (Fig. S.9A) we calculated the mean pairwise squared allele-frequency correlation (r^2) for mapped markers. A linkage disequilibrium

window of 50 (providing > 3 million pairwise values of r^2) resulted in a mean pairwise r^2 of 0.0979, confirming a low overall level of linkage disequilibrium in *B. oleracea*.

2.3.7 Analysis of seed oil traits aided validation of the associative transcriptomic pipeline

Seed oil trait data collected at Rothamsted Research was used to validate the associative transcriptomics pipeline. Unlike flowering time, seed oil traits are regulated by a relatively simple genetic network, which is robust to environmental changes (Barker *et al.*, 2007; Havlickova *et al.*, 2018). This makes seed oil traits a good control, as there are distinct genetic markers which should be identified if the pipeline is performing efficiently. The percentage seed oil composition for 16 different fatty acids were run through the pipeline, these were palmitic, palmitoleic, stearic, oleic, linoleic, linolenic, arachidic, eicosenoic, behenic, erucic, lignoceric and nervonic acids. Within Brassicas, the oil is characterized by erucic acid content, which generally makes up around 50% of the total fatty acids within Brassica seed. Erucic acid is absent in any other commercial plant oils. Other fatty acids are also highly desirable for particular industries, such as behenic and arachidic acids for the oleochemical industry (Cartea *et al.*, 2019), due to this, seed oil composition is well documented in the literature.

Comparison of SNP associations with seed oil composition, to physical positions of known genes involved in fatty acid biosynthesis added confidence to the pipeline's ability (Fig. 2.14). For palmitic acid content (16:0), one significantly associated marker ($P < 0.05$) was identified on C05 13 gene models from an orthologue of AT1G32200.2 *ACT1*, a known phospholipid acyltransferase family protein in *A. thaliana*. A secondary, non-significant peak can also be seen on C06, in the region of an orthologue of AT1G47578.1, C06 *LIP2P2*, an octanoyltransferase involved in fatty acid synthesis in *A. thaliana.*

For palmitoleic acid content (16:1), a peak was seen seven gene models from C05 *ACT1*, and also for an orthologue of AT3G55360.1 C04, *Enoyl-CoA reductase*, involved in long chain fatty acid elongation reactions that are required for cuticular wax, storage lipid and sphingolipid metabolism in *A. thaliana*. Finally, for nervonic acid (24:1) an association was found one gene model from an orthologue of AT1G08510.1, C05 *Fatty Acy-Triesterase B* (*FATB*), which encodes an acyl-acyl carrier involved in saturated fatty acid synthesis in *A. thaliana*.

Furthermore, association of gene expression levels to seed oil traits was also investigated. GEM association peaks were detected for seed oil composition (Fig. 2.15). 27 significant associations (FDR, $P < 0.05$) were identified for stearic acid (18:0) content. Amongst these was an orthologue of AT4G27030.1, C01 *FAD4*, a fatty acid desaturase known in *A. thaliana* to encode a palmitate desaturase. A significant association was also seen with an orthologue of AT1G51040.1, a protein kinase superfamily protein known in *A. thaliana* to be involved in phosphatidylinositol phosphorylation. For oleic acid content (18:1), 235 significant associations were identified (FDR, $P < 0.05$). An orthologue of AT4G34520.1, C01 *FAE1*. FAE, also known as 3-Ketoacyl-CoA Synthase 18, was identified. This is known to be involved in the biosynthesis of long chain fatty acids in both *A. thaliana* and Brassica. Significant associations were also seen with an orthologue of AT5G03860.2, C02 *Malate Synthase* (*MLS*) and AT3G58150.1, C03 *OPA3*, both known to be involved in fatty acid biosynthesis in *A. thaliana*.

Finally, eight significant associations (FDR, $P < 0.05$) were identified for lignoceric acid content (24:0). These eight, included an orthologue of AT5G48880.1 C06 *PKT2*. In *A. thaliana* this gene is known to encode a precursor of 3-Ketoacyl-CoA Thiolase 2, which is involved in the oxidation of fatty acids. Identification of strong candidates for the seed oil traits investigated gave confidence in the pipeline and help to validate its efficacy.

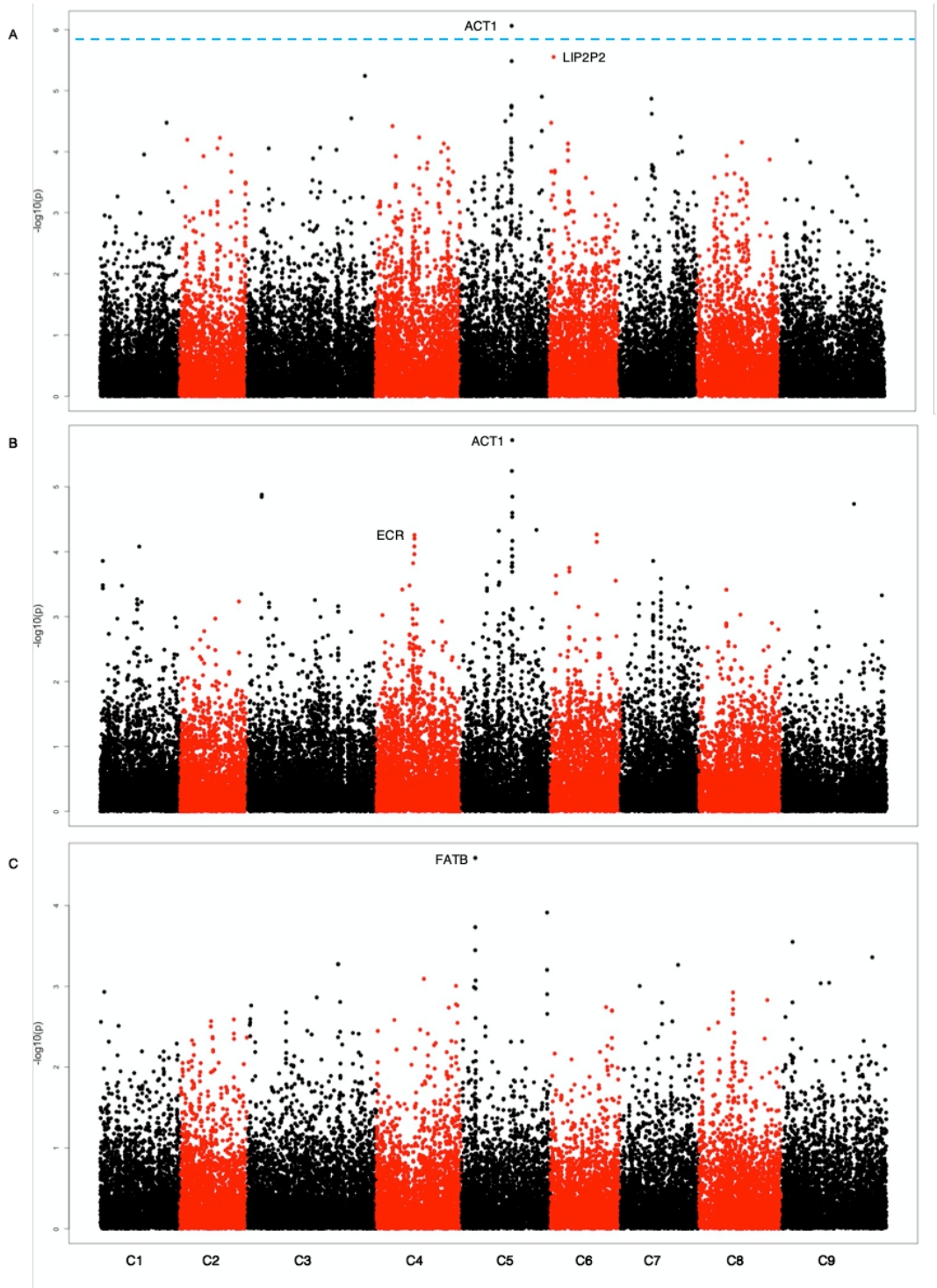


Figure 2.14: Distribution of mapped markers associating with seed oil content. Sixty-nine accessions of *B. oleracea* were phenotyped for seed oil content and marker associations were calculated using a generalized linear model implemented in TASSEL to incorporate population structure. $Log_{10}P$ values were plotted against the nine *B. oleracea* chromosomes in SNP order. A) Palmitic Acid 16:0 B) Palmitoleic Acid 16:1 C) Nervonic Acid 24:1. Blue line FDR threshold, $P < 0.05$.

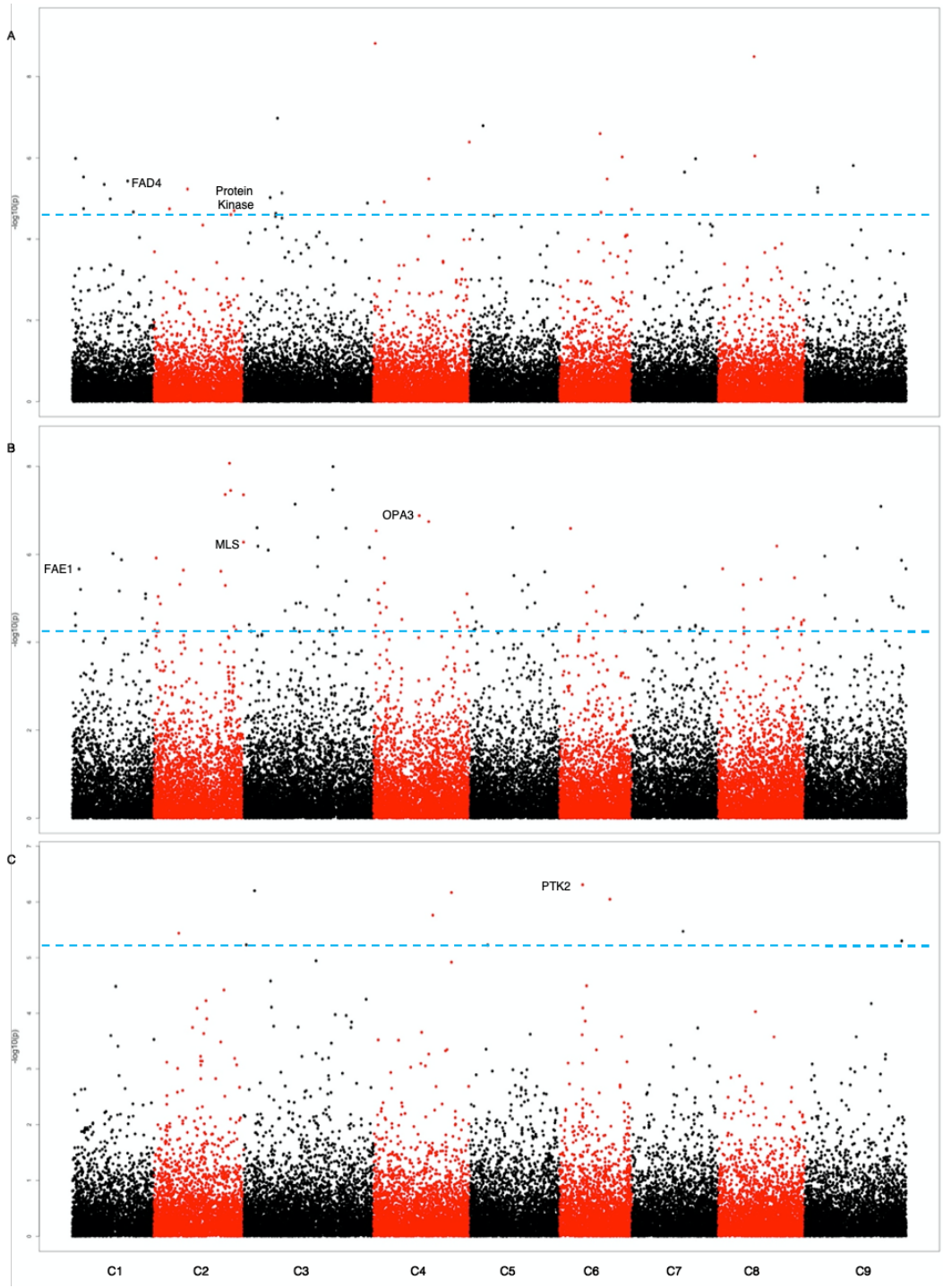


Figure 2.15: Distribution of gene expression markers associating with seed oil content. $Log_{10}P$ values were plotted against the nine *B. oleracea* chromosomes in SNP order. A) Stearic Acid 18:0, B) Oleic Acid 18:1 C) Lignoceric Acid 24:0. Blue line FDR threshold, $P < 0.05$.

2.3.8 *miR172D* identified as a candidate for the control of vernalisation response using associate transcriptomics

SNP associations were compared to the physical positions of orthologues of genes known to be involved in the floral transition in Arabidopsis (Fig. 2.16). A total of 43 flowering time related traits were run through the pipeline. The traits included simple traits such as the days to BBCH51 and BBCH60 for each treatment. Comparative traits were also run, and these included the range in BBCH51 and BBCH60, as well as the differences in days to BBCH51 and BBCH60 within treatments. A total of 111 significant SNPs were identified, $P < 0.05$, six of which were significant associations (Table2.3).

Initially trait data for the non-vernalised experiment was analysed. No significant associations were identified for days to BBCH51, however an association for days to BBCH60 under non-vernalising conditions was identified at Bo8g089990.1:453:T (P = 2.29E-06). This marker was in a region of good synteny to *A. thaliana*, despite there being several unannotated gene models present. Conservation between *A. thaliana* and *B. oleracea* suggest that this region contains an orthologue of *A. thaliana miRNA172D*, AT3G55512 (Fig. 2.16A). *miR172D* has been linked to the floral transition within *A. thaliana* (Aukerman and Sakai, 2003; Poethig, 2009). Furthermore, the difference in days to BBCH51 between six and twelve weeks of vernalisation at 15 ºC, identified a significant association on C07 at Bo7g104810.1:204:T (FDR, $P < 0.05$). This association was in the vicinity of a second orthologue of *miR172D* (Fig. 2.16C).

Table2.3: Significant SNP associations with vernalisation response in diverse *B. oleracea* accessions, detected across the genome (FDR < 0.05), including model information.

Marker Information			Association Information					Model Information	
Marker	**Chromosome**	**Alleles**	**-Log10(p)**	**Marker R^2**	**Traits**	**Arabidopsis ID**	**Orthologue**	**Model**	**Population Structure Correction**
Bo6g103650.1:2010:T	C06	C/T/Y	6.4017787	0.39231	6P 12V 10 ºC DTB	AT1G67140.3	*SWEETIE*	GLM	Q-Matrix
Bo9g179000.1:2589:G	C09	G/T/K	6.4077566	0.39662	6P 12V 10 ºC DTB	AT5G04240.1	*ELF6*	GLM	Q-Matrix
Bo1g011280.1:786:A	C01	A/T/W	6.0844894	0.44220	10P 12V 5 ºC DTF	AT4G31490.1	Coatomer, beta subunit	GLM	Q-Matrix
Bo7g026810.1:124:G	C07	A/G/R	4.7781947	0.36476	6P 12V 10 ºC DTF	AT2G05790.1	O-Glycosyl hydrolases family 17 protein	GLM	PCA
Bo7g104810.1:204:T	C07	A/T/W	5.9788107	0.41678	10P 6V 15 - 5 ºC DTB	AT3G55512	mir172D	GLM	Q-Matrix
Bo2g009460.1:894:T	C02	C/T	7.6880767	0.40565	10P 6V 5 ºC DTF - DTB	AT5G10140.4	*FLC.C2*	GLM	Q-Matrix

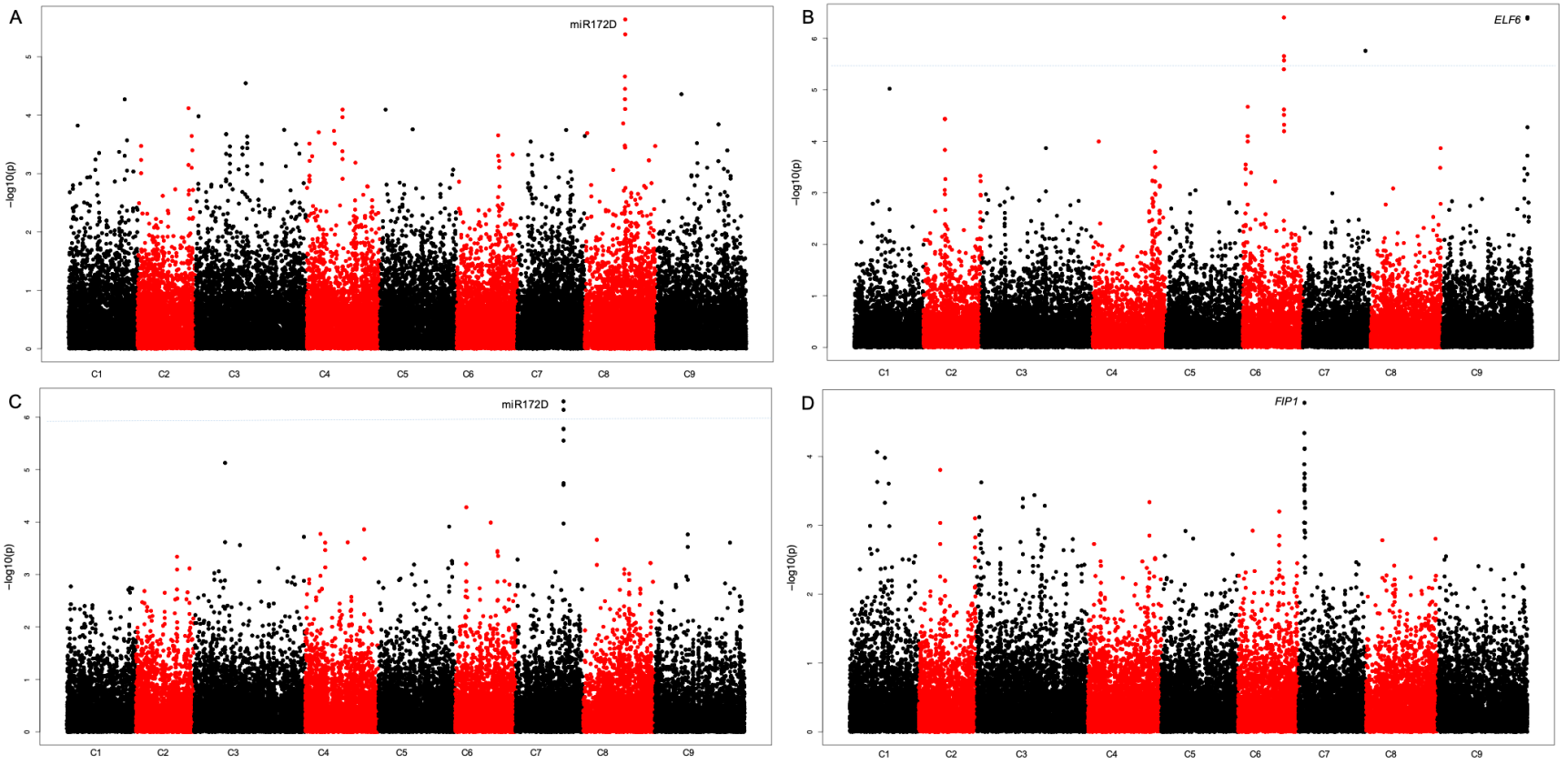


Figure 2.16: The developed pipeline identifies associations with flowering traits. Distribution of mapped markers associating with A) Number of days to BBCH60 under non-vernalising conditions B) BBCH51 after a six-week pre-growth, twelve weeks vernalisation 10 ºC C) The difference in BBCH51 between six and twelve weeks of vernalisation at 15 ºC, after exposure to a ten-week pre-growth D) The BBCH60 after exposure to six-week pre-growth, twelve weeks vernalisation 10 ºC. Sixty-nine accessions of *B. oleracea* were phenotyped for BBCH51 and BBCH60 and marker associations were calculated using a generalised linear model, implemented in TASSEL to incorporate population structure. Log_{10} (P values) were plotted against the nine *B. oleracea* chromosomes in SNP order. Blue line FDR threshold, P< 0.05, FDR

Subsequently, traits relating to the timing of vernalisation were assessed. No significant associations were identified for the six-week pre-growth followed by 5 ºC vernalisation. However, a strong association was identified at the marker Bo7g026810.1:124:G, for the days to BBCH60 following a six-week pre-growth period, with twelve weeks of vernalisation at 10ºC. This was eight gene models from an orthologue of *A. thaliana* C07 *FRI INTERACTING PROTEIN 1,* (*FIP1*), AT2G06005.1 (Fig. 2.16B). *FRI* is a major determinant of natural variation in flowering time in Arabidopsis. Significant associations (FDR, $P < 0.05$), were also identified for the days to BBCH51 following six-week pre-growth and twelve weeks of vernalisation at 10ºC. An association was identified at Bo9g179000.1:2589:G, which is in the vicinity of an orthologue of *Early Flowering 6* (*ELF6*), AT5G04240.1 (Fig. 2.16C). *ELF6* is a nuclear targeted protein able to affect flowering time regulation irrespective of *FLC*.

The differences in flowering phenotype between the SNP variants for the four strongest associations were analysed (Fig. 2.17). There were significant differences in the traits associated with *miR172D* (BBCH60 with no vernalisation and the difference in BBCH51 for plants grown under vernalising temperatures of 5 and 15 ºC) for different alleles (Fig. 2.17A and B). Five individuals contained the A variant for the candidate identified through an association to the difference in BBCH51 after exposure to vernalisation temperatures of 5 and 15 ºC, four of which were broccoli and one cauliflower. The alternate variant was a T allele and was present in 50 individuals. Conversely, the *miR172D* candidate identified through an association with BBCH60 with no vernalisation, had 11 individuals with a C at this locus, whilst 51 had a T. Interestingly, the smaller set with the C allele still contained every crop type.

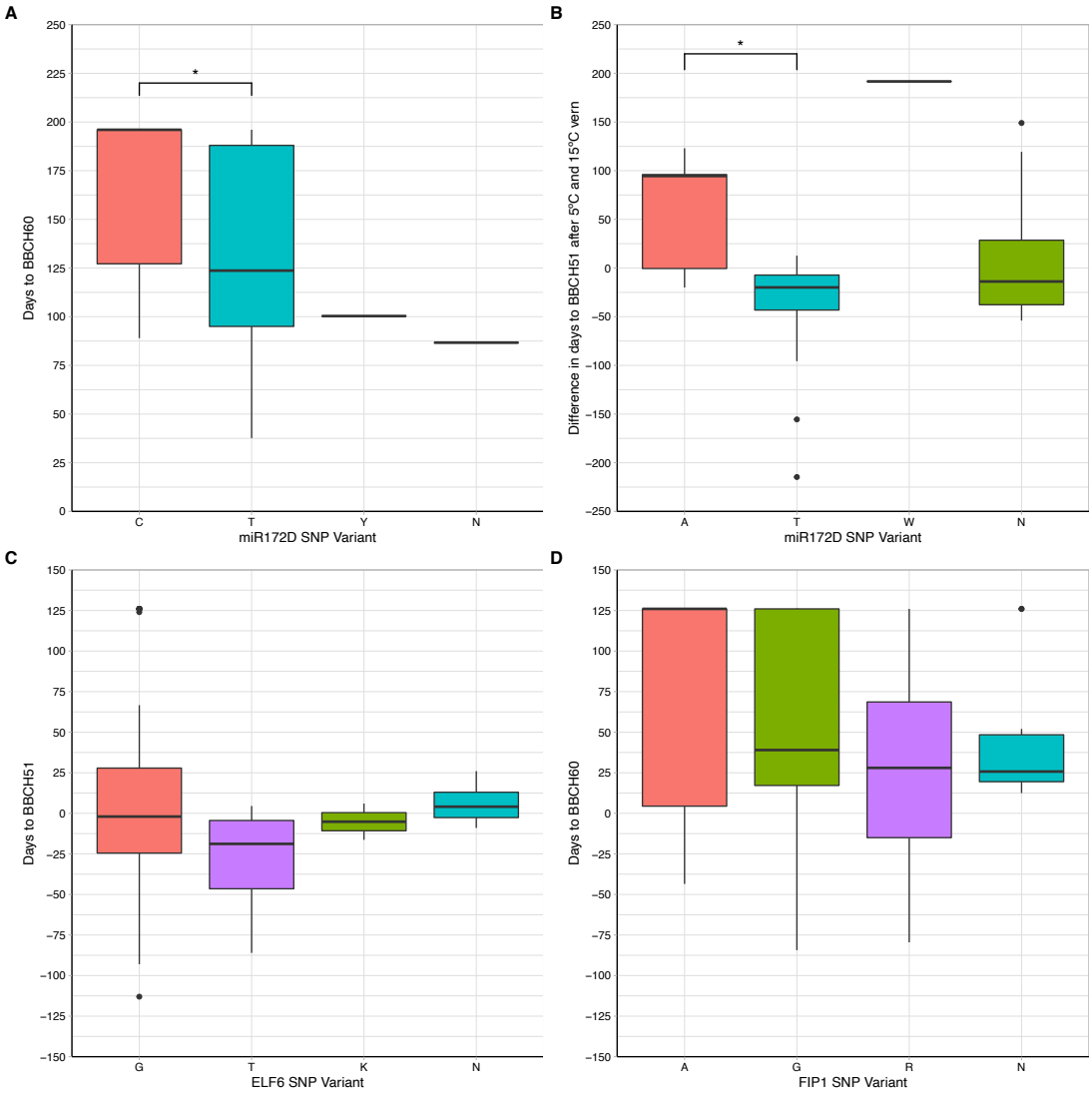


Figure 2.17: A significant phenotypic difference was found for individuals exhibiting SNP variants for the associations pointing to *miR172D* as a candidate.

Boxplots represent the trait data, BBCH51 or BBCH60 for each of the significant markers alongside the different alleles present across the population for each marker. The box represents interquartile range, outliers are represented by black dots.

2.3.9 *BoFLC.C2* identified as a candidate gene for the vegetative-to-floral transition using GEM analysis

To investigate potential links between gene expression and the traits of interest, we performed gene expression marker (GEM) analysis. *BoFLC.C2* was identified as being significantly associated with both the number of days to BBCH51 and BBCH60 under non-vernalising conditions (Fig. 2.18). This is a strong candidate as the role of *BoFLC.C2* in vernalisation requirement is well documented in *B. oleracea* (Irwin *et al.*, 2016). As expected, all five rapid cycling accessions within the population demonstrated no Bo*FLC.C2* expression, a major marker for their lack of vernalisation requirement. A Brassica consortium recently developed targeted sequence capture for a set of genes, including *FLC*. DNA from four of the five rapid cycling accessions had been enriched in that capture library and sequenced. Lacking a reference sequence for *B. oleracea* that contains *BoFLC.C2*, we used *B. napus* (cv. Darmor) (Chalhoub *et al.*, 2014) as a reference to map the captured sequence data from the four rapid cycling accessions to. A 99.54 % identity in coding sequence was revealed when comparing *B. oleracea* transcript data (Irwin *et al.*, 2016) to the Darmor genome reference, justifying the use of the Darmor sequence as a reference. We found that all four rapid cycling accessions, GT050381, GT080767, GT100067 and GT110222, do not have *BoFLC.C2*, which was revealed by a lack of read mapping (Fig. S10A). We investigated mapping for 49 non-rapid cycling accessions where we expect *BoFLC.C2* to be present. For all 49 we found the expected read mapping evidence, confirming that use of the polyploid *B. napus* reference is appropriate (Fig. S10A). It is unlikely that one gene is able to explain all variation over the entire dataset for a complex trait, and indeed only a weak positive correlation (BBCH51 $R^2 = 0.024$, BBCH60 $R^2 = 0.036$) between flowering phenotype and *BoFLC.C2* expression was identified. However, a strong positive correlation (BBCH51 $R^2 = 0.871$, BBCH60 $R^2 = 0.891$) was found for the phenotypic extremes (the rapid cyclers with no expression and the later flowering lines with high levels of *BoFLC.C2*), Fig. 2.19, confirming a role for *BoFLC.C2* in conferring vernalisation response in *B. oleracea*.

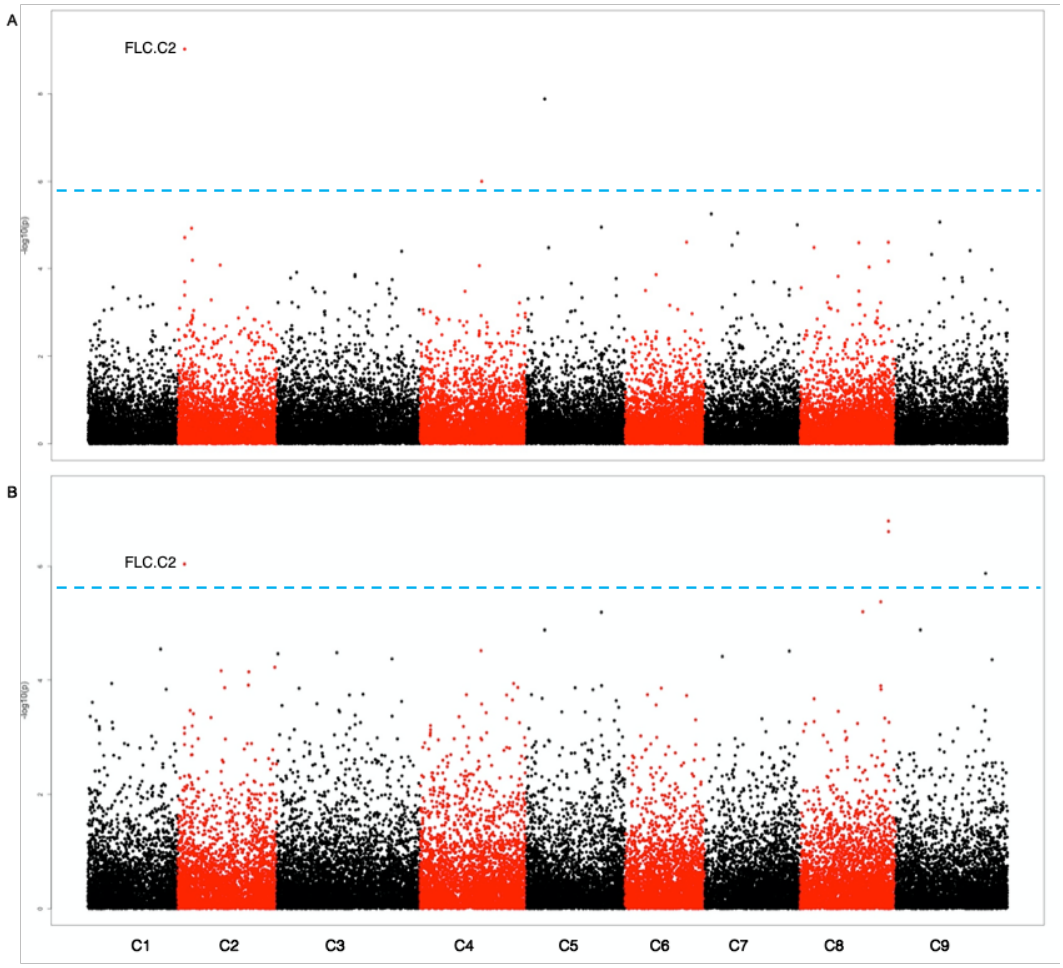


Figure 2.18: Distribution of gene expression markers associating with traits of interest. A) Number of days to BBCH51 after exposure to non-vernalising conditions B) Number of days to BBCH60 after exposure to non-vernalising conditions. $Log_{10}P$ values were plotted against the nine *B. oleracea* chromosomes in SNP order. Blue line FDR threshold, $P < 0.05$.

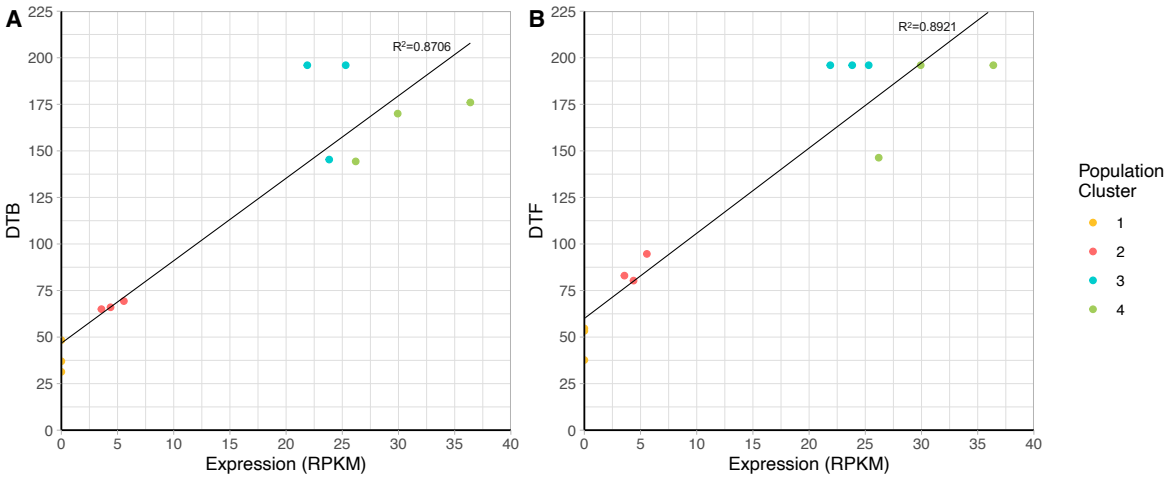


Figure 2.19: A strong positive correlation can be seen between lines at the phenotypic extremes and their *BoFLC.C2* expression levels.
Colours represent the subpopulation of each line, as determined by population structure analysis.

2.4 Discussion

Producing synchronous *B. oleracea* vegetables is a key goal for growers and breeders. Quantifying vernalisation responses for different varieties is an important step towards this goal, providing a foundation for targeted breeding. Without available temperature data, it was important to find a way to assess the comparability of the different experiments to evaluate the effects of environmental perturbations on the vegetative-to-floral transition. Leaf numbers were used as a proxy for developmental stage prior to vernalisation. No significant difference was seen between the numbers of leaves in experiment 1, 3 or under the control conditions, but significantly less leaves were seen for plants grown in experiment 2. This was expected as the plants were all 70 d old when leaf numbers were counted, except for those in sowing 2 which were 42 d old. This indicates that the plants were at a similar developmental stage on entering vernalisation and that the staggered sowing did not affect the ability to compare between experiments 1, 3 and NV.

To assess the effect of the timing of vernalisation, data from experiments 1 and 2 were compared. *A. thaliana* goes through a juvenile phase, distinguished by both morphological and genetic factors. Morphological in the form of leaf shape and genetic in the form of variation in levels of *miRNA156*, and whilst it has been documented in the literature that *B. oleracea* has a juvenile phase (Hand and Atherton, 1987; Booij and Struik, 1990; Walley *et*

al., 2012; Wurr *et al.*, 2016), no such morphological or genetic markers have been identified. To identify lines which demonstrated a juvenile phase and subsequently begin searching for markers for juvenility, plants were exposed to two different pre-growth lengths. Results mirrored what was predicted, with more than half of accessions flowering later following the six-week pre-growth period, suggesting the plants had to continue to grow during vernalisation to reach their adult phase and respond to floral inductive cues.

Of the 47 lines which reached BBCH51 earlier following the ten-week pre-growth, only seven were significantly earlier with all not reaching BBCH51 during the experiment following the six-week vernalisation treatment. These phenotypes suggest these lines require the extra four weeks pre-growth to reach their adult phase of growth and be able to respond to the vernalisation treatment as a floral inductive cue. The juvenile phase is a complex problem, the vernalisation requirement of a plant must be known and fully saturated before true investigation of juvenility can occur, as the two are interlinked. Flowering cannot occur if the plant is in the juvenile phase, or if the vernalisation requirement has not been met. Identifying the optimal vernalisation length and temperature for candidate lines identified here and investigation with an increased range of pre-growth lengths would help to provide more detailed information on the juvenile phase within these lines.

Vernalisation temperature and length are known to be key determinants of the timing of the floral transition (Putterill, Laurie and Macknight, 2004; Rosen *et al.*, 2018). It is thought the two factors are intrinsically linked in meeting the vernalisation requirement of a plant. The effects of vernalisation temperature and length were not immediately obvious, a high-level look at the data revealed that similar numbers of individuals reached BBCH51 between the three temperature treatments and between the two vernalisation lengths. Furthermore, whilst the mean days to BBCH51 was significantly different between the two vernalisation lengths, between the temperatures only the difference between mean days to BBCH51 for the 5 and 10 ºC treatments was significant. However, assessment of the distribution in days to BBCH51 and BBCH60 revealed that increased vernalisation length and decreased vernalisation temperature, resulted on averaged in faster and more synchronous heading and flowering within this population.

Synchronous heading is a very desirable trait for growers allowing for an easier and higher yielding harvest. With the prospect of warmer winters as a result of climate change, achieving a sufficient length of vernalisation at a cool enough temperature to achieve synchronous heading is growing more unlikely for some accessions of *B. oleracea*. Consequently, for those crop types which are reliant on the floral transition, it is important to identify lines which are

able to transition synchronously under warmer and shorter periods of vernalisation. Within the phenotyped population, many lines were identified with a low or non-existent vernalisation requirement, for example the case study of GT110244 (Fig. 2.11) which demonstrates faster development under warmer vernalisation temperatures. The non-vernalised treatment enabled identification of 47 lines which do not require vernalisation to be able to transition through to the floral state. However, it is not just the ability to pass through this floral transition under these warmer conditions which is important, but the ability to do so and maintain desirable phenotypes. Many of the lines which did not require vernalisation to undergo the floral transition, lost their desirable phenotype. As a consequence of this, whilst these lines would not be suitable as direct alternatives to the commercial lines we use today under warmer conditions, they could provide key genetic markers that could confer a lack of vernalisation requirement. These genetic markers could provide the basis for future breeding efforts in the face of climate change.

In order to ascertain such gene candidates involved in vernalisation response, associative transcriptomics was run on the flowering trait data obtained from the phenotyping, at the University of York (L. Havlickova). Flowering is known to be a complex trait, heavily influenced by environmental factors combined with a large number of genes involved in its regulation. These qualities can make it difficult to identify strong candidate leads using this type of analysis. However, numerous GWAS analyses have been carried out on flowering time and yielded robust results (Huang *et al.*, 2012; Xu *et al.*, 2015; Romero Navarro *et al.*, 2017; Shah *et al.*, 2018; Prom *et al.*, 2019; Raman *et al.*, 2019). No significant associations were identified on this first analysis of the data. To test the pipeline's efficiency, seed oil trait data obtained from Rothamsted Research for the population was analysed using associative transcriptomics. Genes controlling oil traits are well documented and known to be identified using this type of analysis (Gacek *et al.*, 2017; Havlickova *et al.*, 2018; Gazave *et al.*, 2020), however in this instance they were not. The QQ plots for the data showed clear under or overcorrection and many had breaks in. This resulted in the presence horizontal lines of markers through many of the Manhattan plots. As a result, it was predicted that population structure may have been causing an issue and was subsequently investigated.

Population structure is extremely important in GWAS analysis. If exact, accounting for population structure ensures markers are only identified due to a significant association with the trait of interest and not due to relatedness between individuals in the population. To ensure the population structure was representative, a stringent set of defined criteria was used to select unlinked markers, from which to generate a novel population structure specific to the panel of lines used here. The population size for GWAS experiments is often designed to be as large

as possible to encompass the maximum amount of variation, however here the population was limited to just the 69 lines phenotyped. Despite this, previously it has been demonstrated that meaningful results can be obtained from GWAS using less than 100 accessions (Atwell *et al.*, 2010). Here we evaluated phenotypic and crop type information to assess the reliability of the subpopulations identified and then further used phylogenetic analysis to validate this through clustering. Due to the small population size, it was important to validate the population structure, and this was done through analysis of the seed oil trait data through the pipeline as a control. Running the control data enabled identification of known candidate genes involved in fatty acid biosynthesis. This positive result demonstrated that the new population structure was effective and subsequently the flowering trait data was analysed using the updated population data.

Using this validated population structure with associative mapping, candidates orthologous to known *A. thaliana* floral regulators were identified. GWAS analysis found *miR172D* to be significantly associated to two flowering traits (Fig. 2.16). In *A. thaliana*, the *miR172* family post-transcriptionally supress a number of *APETALA1*-like genes, including *TOE-1, 2* and *3*, which in turn aids the promotion of floral induction (Jung *et al.*, no date; Aukerman and Sakai, 2003; Wu *et al.*, 2009; Teotia and Tang, 2015). Furthermore, the SNP variant data for both associations that point to *miR172D*, exhibit significant phenotypic differences. Two orthologues of *A. thaliana miR172D* have currently been identified in *B. oleracea* (Shivaraj *et al.*, 2014) but their functional roles have yet to be determined.

Whilst some candidate genes have been identified, many more significant associations were detected but could not be identified, as they fell in unannotated regions of the reference sequence. The reference sequence used was the C file of the *B. napus* pantranscriptome (He *et al.*, 2015). This is constructed of the coding DNA sequence gene models from the *B. oleracea* TO1000 genome sequence assembly (Parkin *et al.*, 2014). TO1000 was included in this study, it is a rapid cycling *B. oleracea* - a cross between a Chinese kale and a broccoli. It's rapid cycling nature means it requires no vernalisation to flower, therefore it will be genotypically different to the later flowering lines which confer a strong vernalisation requirement. This combined with the fact that the SNP data was obtained from transcriptome data of leaf tissue at one time point only, means that many SNPs may have been excluded from this analysis; if they are not expressed in the leaf or if they are non-coding. Use of transcriptome data from other tissues, such as apex tissue, which is known to be important during the floral transition, could enable the identification of a greater number of associations.

GEM analysis identified a significant association with Bo*FLC.C2* with both BBCH51 and BBCH60 under non-vernalising conditions. This can be attributed to lines at the phenotypic extremes of the population (Fig. 2.19). No Bo*FLC.C2* expression was detected in five lines. A loss-of-function mutation at Bo*FLC.C2* in cauliflower has been associated with an early flowering phenotype (Ridge *et al.*, 2015), indicating that Bo*FLC.C2* has an equivalent role in cauliflower to *FLC* in *A. thaliana*. Bo*FLC.C2* exhibits very similar genomic organisation to At*FLC*, with seven exons of similar size (Lin *et al.*, 2005) and is known to control requirement for vernalisation in *B. oleracea* (Irwin *et al.*, 2016). According to the EcoTILLING data analysed, four of the five lines for which Bo*FLC.C2* expression could not be detected, did not have the Bo*FLC.C2* paralogue. All of these lines were rapid cycling kales and demonstrated an early flowering phenotype, suggesting that Bo*FLC.C2* has a similar role to At*FLC* also in kales, and potentially across *B. oleracea*. Although BBCH51 and BBCH60 were highly correlated with Bo*FLC.C2* expression under non-vernalising conditions for the phenotypic extremes, for the whole population the correlation was low. This is to be expected as flowering is known to be a complex trait that requires a combination of environmental cues and the coordinated expression of many gene regulators, therefore Bo*FLC.C2* is unlikely to account for all the observed variation.

Expression data used for the GEM analysis was generated from leaf tissue at one timepoint, consequently any genes which are not expressed in that tissue at that time will not be identified in this analysis. Looking forward, use of other tissues at later timepoints to generate transcriptome data could provide a means to identify a greater number of associations.

2.5 Conclusion

Here, a novel associative transcriptomics pipeline has been generated and validated for *B. oleracea*. Furthermore, it has been demonstrated that this pipeline is effective in identifying genetic regulators of complex traits, such a flowering time. The ability to identify genes underlying phenotypic traits in *B. oleracea* is an important step for the improvement of brassica vegetables and can provide a foundation for breeding programmes. GWAS analysis identified *miR172D* as a candidate for vernalisation response, whilst GEM analysis identified a significant marker at *BoFLC.C2*, an important gene in the vernalisation pathway of *B. oleracea*. These results provide insight into the genetic control of flowering in *B. oleracea*, information that could be used in future breeding strategies in light of changing climatic conditions. Further work looking at a wider range of treatments and using more tissues,

specifically apex tissue, to generate expression data could result in the identification of a more comprehensive list of gene candidates that could be used in further analysis. The identification of gene candidates is just the first step, once identified further analysis could be carried out in the form of generating knockouts of these genes within *B. oleracea* or performing complementation assays within *A. thaliana*, to assess gene function.

Chapter 3

Investigating the genetic basis of the floral transition in DH1012 using transcriptomics

3.1 Introduction

3.1.1 The leaf and apex play distinct roles in the floral transition in Arabidopsis

Both the leaf and apex tissues have been demonstrated to play distinct roles in the floral transition in Arabidopsis. The leaf has a key role in initiating flowering through its ability to sense photoperiod (Kardailsky *et al.*, 1999; Abe *et al.*, 2005; Romera-Branchat, Andrés and Coupland, 2014). Photoperiod is a key signal in the floral transition, enabling plants to align their flowering to favourable seasons. *CO* is expressed in response to long days and its expression in the leaf induces the expression of the floral promoter *FT* (Yoo *et al.*, 2005). *FT* is a long-distance flowering signal that is expressed in the vasculature of cotyledons and leaves (Abe *et al.*, 2005). Once *FT* reaches the shoot apical meristem it triggers a host of downstream gene targets that act to induce the transition from the vegetative to floral phase. These downstream targets include the floral promoters *FUL, SOC1* and *AP1* (Kardailsky *et al.*, 1999).

The role of *FT* has also been studied in other members of the *Brassicaceae* family. *B. napus* contains six *FT* paralogues. Through the use of EMS mutants, it was demonstrated that despite the redundancy in the *B. napus* genome, point mutations in two of these six paralogues, *BnC6.FTa* and *BnC6.FTb,* led to an effect on flowering (Guo *et al.*, 2014), suggesting there is more than one paralogue which plays a role in the floral transition. Furthermore, through GWAS analysis in *B. napus*, it has been suggested that *FT* paralogues are also involved in the modulation of yield-related traits in addition to flowering time (Raman *et al.*, 2019) and a similar conclusion was made for *FT* paralogues in *B. juncea*. The overexpression of *BjuFT* in Arabidopsis *ft-10* mutants, implicated *BjuFT* in multiple traits beyond flowering, including lateral branching, silique shape, seed size, oil-profile, stomatal

morphology and plant height (Tyagi *et al.*, 2018). In contrast to this, in a rapid cycling *B. rapa* line Sarisha-14, *FT* does not seem to be a key player in the floral transition and as of yet no additional roles for *FT* have been documented. *B. rapa* is known to contain two paralogues of *FT, BraFT.A02* and *BraFT.A07,* with *BraFT.A02* known to be the main *FT*-like gene regulating the floral transition (I *et al.*, 2019). Expression of *BraFT.A02* in the leaf was not detectable prior to the floral transition, denoting that flowering in Sarisha-14 occurs independently of *FT* expression levels and instead is reliant on the ageing pathway to trigger its early flowering phenotype (Calderwood, Hepworth, *et al.*, 2021). This work suggests that overall Brassica *FT* copies play a wider role than our current understanding of their Arabidopsis homologue and that in rapid cycling lines *FT* could be excluded to accommodate for an early flowering phenotype.

3.1.2 Transcriptomics facilitates a detailed insight to gene expression

Transcriptomics is the study of the transcriptome, which is the entirety of RNA transcripts produced by the genome, at a specific time in a tissue or cell. Transcriptomics uses high through-put methods such as microarrays or RNA-Sequencing (RNA-Seq). RNA-Seq applies next generation sequencing (NGS) to sequence and allow quantification of transcripts in a biological sample. With advances in NGS technologies, RNA-Seq has become a popular method, particularly for the expression quantification and the identification of differentially expressed genes (DEGs) and has been applied to the study of many developmental transitions, including the floral transition in a range of species (Gao *et al.*, 2014; Huang *et al.*, 2014; Klepikova *et al.*, 2016; Hickman *et al.*, 2017; Braynen *et al.*, 2021; Kim *et al.*, 2021; Xanthopoulou *et al.*, 2021).

Transcriptome sequencing in Arabidopsis has been used to investigate the floral transition in the apical meristem (Klepikova *et al.*, 2015), leading to the identification of a critical period between 10 and 12 d post sowing. During this period *FLC* expression is reduced and *LFY* expression had yet to increase, and this was characterised by an increase in DEG number. Further investigation of this DEG peak through gene ontology (GO) enrichment analysis, revealed that this was due to overrepresentation of genes involved in the cell cycle. This suggested that during this critical period in Arabidopsis, an acceleration is seen in the rate of cell divisions, and this partially synchronises with cell cycling.

As well as Arabidopsis, transcriptome-wide analyses have been applied to other members of the *Brassicaceae* family. RNA-Seq was used to generate a gene regulatory network to describe flowering in Chinese cabbage (*Brassica campestris*) (Huang *et al.*, no date). The stalk is the

harvestable product for Chinese cabbage and as such yield is directly related to stalk development. In this study, RNA-Seq of apical meristem tissue was used to investigate stalk development at three key stages: seedling, bolting and flowering, to better understand the genetic regulators of stalk development. 11514 DEGs were identified among these three stages of stalk development and through GO analysis, enrichment was found for the 'plant hormone signal transduction' and 'ribosome' pathways. From the data obtained, a putative gene network was constructed using the four major floral pathways of the Arabidopsis flowering time network: autonomous, vernalisation, photoperiod, and gibberellic acid, based on knowledge from Arabidopsis. However, despite the presence of multiple gene copies within *B. campestris*, paralogues were grouped together for the purpose of generating the gene network.

Whole transcriptome analysis has also been carried out in another member of the *Brassicaceae* family, *B. napus* (Shah *et al.*, 2018). Oilseed rape (*B. napus*) is globally one of the most important sources of vegetable oil and through modification of flowering time has been adapted to grow in many geographical regions. RNA-Seq was carried out on the semi-winter variety Ningyou7, to better understand the floral transition. Through combining RNA-Seq analysis with association analysis and genetic mapping, this study identified 36 genes with an association to flowering time and identified potential neofunctionalization in homologues of known flowering time genes including *FUL* and *VIN3*.

A further study in Lychee (*Litchi chinensis*) applied RNA-Seq to the analysis of apical meristems to reconstruct the regulatory network underlying the transition from vegetative to floral development (Lu *et al.*, 2017). Lychee is an important crop in Southern Asia, in which the floral transition in induced by low temperatures and enhanced by drought. However, warm winters, which may be a result of global warming, are interfering with this transition, leading to reduced yields as a result of poor flowering. Therefore, similar to many *B. oleracea* cultivars, changing climatic conditions is posing a threat to the need for cold for flowering. Finding a way to bypass this need is crucial to the Lychee industry, and an understanding of the genetics of this transition may be a way to do that. Previous studies had demonstrated that reactive oxygen species (ROS) could promote flowering by increasing LcAP1 expression at 18/13 °C (day/night), temperatures which are too high to induce the floral transition in Lychee, thus suggesting ROS may be able to partially bypass the chilling effect (Zhou *et al.*, 2012). As a result of this five-year-old Lychee trees were grown at a low (15/8 °C), medium (18/13 °C) or high (25/20 °C) temperature treatments. An additional category of trees exposed to medium temperatures and ROS was also included. Apical meristem tissue was sampled at 0 d (pre-floral induction), 30 d (floral induction) and 75 d (floral initiation) for each of the

treatments. RNA-Seq was applied to compare the expression profiles of the apical meristems between treatments. DEGs linked to the low temperature or medium temperature plus ROS treatment were focused on. Unlike the previous studies, this study constructed a novel (species-specific) gene network. The network constructed was large and extremely complex, containing over 400 unigenes. Pearson's correlation coefficient was calculated to obtain positive or negative correlations between genes from which to construct a network (Lu *et al.*, 2017). Clustering the genes revealed four distinct hubs in this network, perhaps therefore a simpler summary network, like that which has been described in Arabidopsis (Jaeger *et al.*, 2013), could be constructed to demonstrate only the key interactions, consequently making the network easier to interpret.

In addition to the use of RNA-Seq to investigate developmental transitions, it has also been used to produce valuable resources, such as the construction of the Brassica A and C genome based pantranscriptome (He *et al.*, 2015). This was designed to be a valuable tool for *B. napus* research, but is also a useful tool for the diploids *B. rapa* and *B. oleracea*.

3.1.3 Hypotheses and Aims

To investigate the genetic basis of the floral transition in *B. oleracea*, a transcriptome time series experiment was designed and carried out. The rapid cycling *B. oleracea* variety DH1012 was selected, and leaf and apex tissues sampled from the plants over time. This chapter discusses this experiment and explores the differences between leaf and apex tissues over the floral transition. It also investigates the expression patterns of orthologues of key Arabidopsis floral regulators in an attempt to uncover more about their roles in the floral transition in *B. oleracea*.

Understanding the floral transition in *B. oleracea* is of great importance for the continued production of synchronous crop vegetables. Much of our current knowledge on flowering stems from the model species *A. thaliana*, in which the leaf and apex tissues play distinct roles in this transition. The floral transition in *B. oleracea* is marked by morphological changes at the apex. Based on the similarities between *A. thaliana* and *B. oleracea*, I hypothesise that there will be differentiation in gene expression between the leaf and apex, following morphological changes at the apex, which will indicate the switch to the floral state.

Here tissue has been sampled across time and the apex monitored morphologically to determine a period over which the floral transition occurred in DH1012. Light microscopy demonstrated that the apex had transitioned to floral state at day 35 of the time series. I

hypothesise enrichment for GO:0009908 flower development, will be seen in both leaf and apex tissues at day 35.

In Arabidopsis, *FT* is an important gene in the photoperiod pathway of the floral transition, yet in rapid cycling *B. rapa* Sarisha14 the role *FT* is less clear and instead the ageing pathway seems to be more important for flowering. Based on these observations, I hypothesise that *FT* will not play a key role in the vegetative-to-floral transition in DH1012 and that instead members of the ageing pathway will be important.

3.2 Methods

3.2.1 Plant Growth and Sampling

Seeds from *B. oleracea* DH1012, a genetically fixed doubled haploid resulting from the cross between a Chinese kale and a broccoli, were sown in cereals mix (40 % Medium Grade Peat, 40 % Sterilised Soil, 20 % Horticultural Grit, 1.3 kg/m³ PG Mix 14-16-18 + Te Base Fertiliser, 1 kg/m³ Osmocote Mini 16-8-11 2 mg + Te 0.02 % B, Wetting Agent, 3 kg/m³ Maglime, 300 g/m³ Exemptor). Plants were grown in Conviron MTPS 144 controlled environment room with Valoya NS1 LED lighting (250 µmol m^{-2} s^{-1}) 18 °C day/15 °C night, 70 % relative humidity with a 16-hr day. First true leaf and apex samples were taken on ice over development during the vegetative growth and the floral transition, continuing until floral buds were visible (developmental stage BBCH51 (Meier, 2001)). Each sample consisted of a single leaf or apex and three replicates were taken at each timepoint and harvested into LN_2 before storage at -70 °C. The apex was monitored over the time series using light microscopy, apices from three plants were taken at the same time as samples for expression. Sampling was ended when 50% of the remaining plants had reached BBCH51 (Meier, 2001), buds visible when looking down on the plant.

For RNA extraction, samples were ground in LN_2 into a fine powder. RNA extraction was performed using the E.ZN.A. Plant RNA Kit (Omega Bio-tek Inc., Norcross, Georgia, http://omegabiotek.com/store) including the optional DNase treatment. RNA samples were sent to Novogene, Beijing, China, where 150 bp paired-end reads were generated. Sequencing was performed using the Illumina HiSeq X platform. An average of 58.31 M reads were generated per sample.

3.2.2 Aligning reads and quantifying expression levels

Quality control was carried out using FastQC (Anders, 2010) on raw sequencing reads. Trimmomatic (Bolger, Lohse and Usadel, 2014) was used to remove the adaptor sequences from the reads and trimmed reads were aligned to the reference sequence using HISAT v2.0.4 (Kim, Langmead and Salzberg, 2015). Reads were aligned to both the *B. oleracea* pantranscriptome (He *et al.*, 2015) and the *B. oleracea* pangenome (Golicz *et al.*, 2016). SAM files outputted from HISAT were sorted and converted to BAM files using SAMTools (Li *et al.*, 2009) and StringTie v1.2.2 (Pertea *et al.*, 2015) was used to quantify gene expression. Gene expression level is reported in Counts per Million (CPM) or Transcripts per Million (TPM). Orthologues of Arabidopsis genes in *B. oleracea* were identified by H. Woolfenden based on sequence similarity and gene synteny using BLAST. Principle component analysis (PCA) was carried out using prcomp() function as part of the stats package (Team, 2013).

3.2.3 Differential gene expression analysis and gene ontology

Differential gene expression analysis was carried out using EdgeR (Robinson, McCarthy and Smyth, 2010). Genes were filtered to remove those with no detectable expression. A logFC threshold of log2(1.5) and an FDR < 0.05 were used to identify DEGs. Gene ontology (GO) enrichment was performed for DEGs identified *(P* < 0.05) using Dicot PLAZA 4.5 (Van Bel *et al.*, 2018). The background gene list consisted of all genes expressed across the time series with a TPM > 1.

3.2.4 Sequence alignment

Sequence alignment was carried out using the MUSCLE sequence alignment tool (Edgar, 2004) within Geneious Prime 2021.2.2 (Kearse *et al.*, 2012). The default parameters were used, including a maximum number of iterations of eight. Sequences for Arabidopsis genes were obtained through TAIR 10 (Huala *et al.*, 2001). Phylogenetic trees were constructed from genomic sequences in Geneious Prime 2021.2.2, using the Neighbour Joining method (Saitou and Nei, 1987).

3.2.5 Self-organising maps

The Kohonen library (Kohonen, 1997) in R was used construct self-organising maps (SOMs). Only DEGs were selected for clustering, thus only genes whose expression varied significantly over time were analysed. Counts per million (CPM) were used and data was

normalised prior to carrying out SOM analysis to have a mean of 0 and standard deviation of 1. For the apex tissue, 8697 genes were clustered into a 3 x 3 toroidal SOM, whilst for the leaf tissue a larger 4 x 4 toroidal SOM was used to cluster the 17243 genes (Kohonen, 1997; Wehrens and Buydens, 2007). 100 training iterations were used during clustering, over which the a-learning rate was decreased from 0.05 to 0.01 (Chitwood, Maloof and Sinha, 2013).

3.3 Results

3.3.1 Experimental design and sample collection

To investigate control mechanisms for the floral transition in *B.* oleracea, leaf and apex tissue were sampled across development. The apical meristem and leaves are known to be tissues in which key floral genes are expressed (Srikanth and Schmid, 2011; Andrés and Coupland, 2012). The leaves are involved in plant primary metabolism and light capture, consequently sampling leaf tissue allows for the study of both the circadian clock genes and photoperiod pathway. To ensure biologically equivalent tissue, the first true leaf (the first leaf to form after the cotyledons) was taken and consequently, leaf samples were only taken until 35 d. This was because at this point the leaves had begun to show signs of senescence and were excluded from further sampling, due to the predicted expression of senescence and stress related genes. Sampling the first true leaf across the time series means the effect of age was introduced into the samples. Consequently it is possible to be able to study the ageing pathway across the time series, which is important as plant age is known to be a factor involved in the induction of flowering (Yang *et al.*, 2013).

The apex is known to be the site of expression for numerous floral integrators and was a necessary tissue to sample to study the floral transition. As much of the surrounding tissue was removed from the apical meristems that were collected and this was carried out by hand using a scalpel, on ice. The apical meristem is responsible for producing all of the above ground organs of the plant. Due to the tissues collected, destructive sampling had to be carried out, as a consequence the apex was monitored to assess the variation within the population and to determine outward signs of the floral transition at the apex. A schematic of the sampling process is displayed in Fig. 3.1

To limit subjectivity in the sampling, the BBCH scale was used to determine the end point of the sampling (Meier, 2001). Sampling finished when over half of the remaining plants reached

BBCH51, which is denoted as when flower buds are visible on the plant from above. For this experiment BBCH51 was reached at 51 d post sowing.

Figure 3.1: Schematic to represent the sampling schedule for collection of transcriptome time series data.
Red numbers displayed below the bottom axis indicate days post sowing at which apex and leaf tissues were sampled, orange numbers represent apex tissue only. The representations of the plants indicate the number of visible leaves on the plants at those timepoints.

3.3.2 Meristem morphology across the floral transition

The apex was monitored across the time series using light microscopy (Fig. 3.2). The development of the shoot apical meristem was monitored across the time series using light microscopy, to determine when the meristem had transitioned from vegetative to inflorescence. At 29 d the apex was still in a vegetative state (Fig. 3.2A); leaf primordia can clearly be seen surrounding the shoot apical meristem. By 35 d rounded floral primordia were seen surrounding the apex (Fig. 3.2B) indicating that the floral transition had occurred at the apex and the plant was producing floral tissues. This demonstrates that the floral transition occurred between 29 d and 35 d in DH1012. The apex was monitored after the floral transition had occurred and at 41 d further development of the floral primordia was seen (Fig. 3.2C), as they began to take on a form more reminiscent of the bud tissue they would ultimately become.

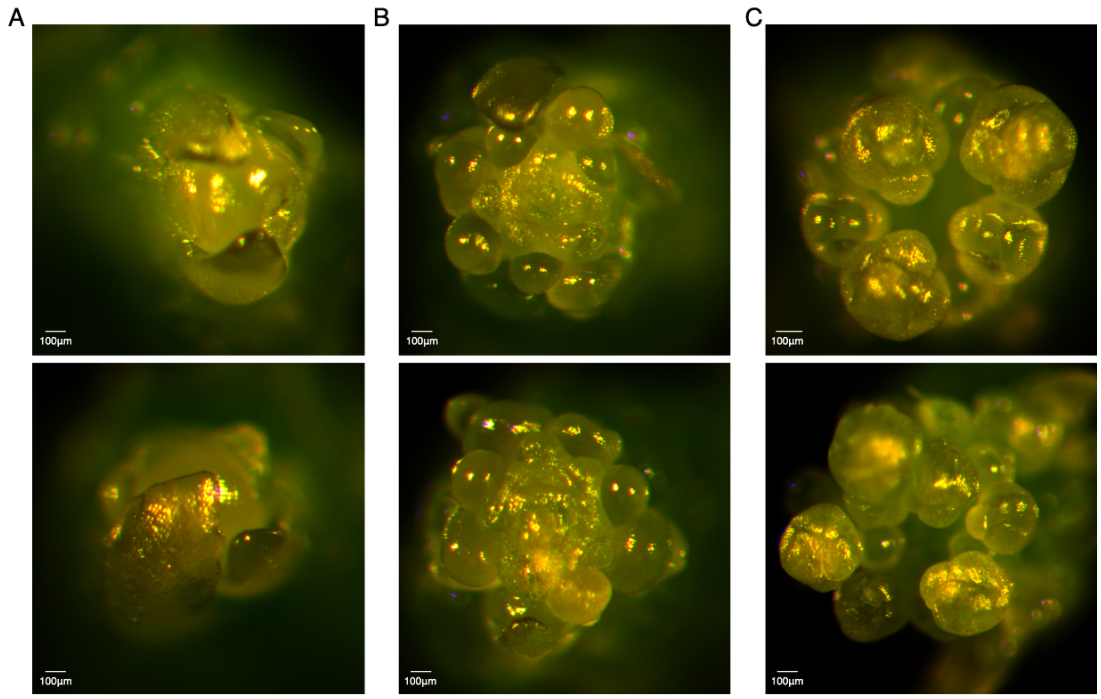


Figure 3.2: The floral transition can be monitored through morphological changes at the shoot apical meristem.

A) 29 d apices, leaf primordia surround the meristem tissue. B) 35 d apices, rounded floral primordia are visible around the meristem tissue. C) 41 d apices, further development of the floral primordia is observed.

3.3.3 Reference genome sequence and gene models

For RNA-Seq analysis, short reads must be aligned to a reference sequence. Two reference sequences were compared for this experiment, the C file of the *B. napus* pantranscriptome (Table S.3B) (He *et al.*, 2015) against the *B. oleracea* pangenome (Table S.4B) (Golicz *et al.*, 2016). Using the pantranscriptome, the number of mapped reads ranged from 28.04 M, with 98.89% of those reads uniquely mapping in sample D12A5_1, to 44.77 M with 99.01% of those reads uniquely mapping, in sample D12L8_2. However, the total number of mapping reads was much higher when using the pangenome sequence, ranging between 47.42 M with 92.4% of those reads uniquely mapping in sample D12A8_1 to 70.13 M with 89.35% of those reads uniquely mapping in sample D12SL_3. Due to the higher mapping rate, the pangenome was selected as the reference for the data, giving an average of 58.31 M mapped reads per sample (SD 6.3) and an average of 91.36% of reads uniquely mapping (SD 0.017).

3.3.4 Principal component analysis

Principal component analysis (PCA) was carried out using TPM values after normalisation. Good separation between the leaf and apex samples was achieved using the first two principal components, PC1 and PC2, which together explained ~ 45 % of the total variance (Fig. 3.3). A greater spread in replicates was observed for the later samples. As a result, PC1-5 were plotted, which resolved an outlier in PC1 and PC3 and PC1 and PC5 (Fig. 3.4). This outlier was 51 d apex sample, replicate 3. The concentration of RNA sent was ~ 600 ng/µl, whilst the read number for this library was not significantly different from the remaining samples at approximately 52.06 M reads, of which 92.81% were uniquely mapped. These results indicate there was not an issue with the sequencing and instead point toward this particular plant being further along in its development than the other two replicates for this timepoint.

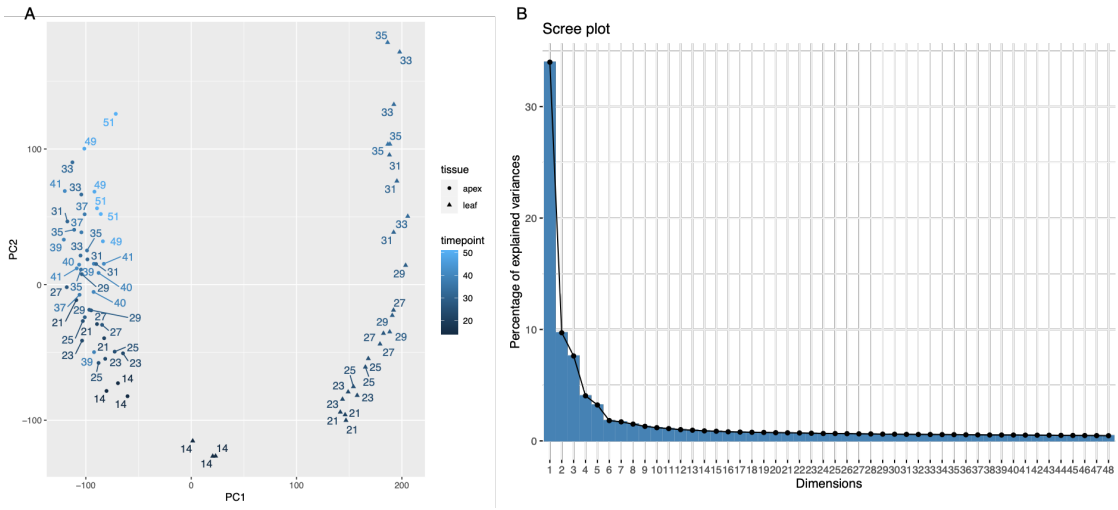


Figure 3.3: Principal component analysis (PCA) data from leaf and apex transcripts in *B. oleracea* DH1012.

A) PCA plot comparing PC1 and PC2. Leaf and apex tissue samples cluster together. B) Scree plot demonstrating ~ 45% of the total variance can be explain by PC1 and PC2.

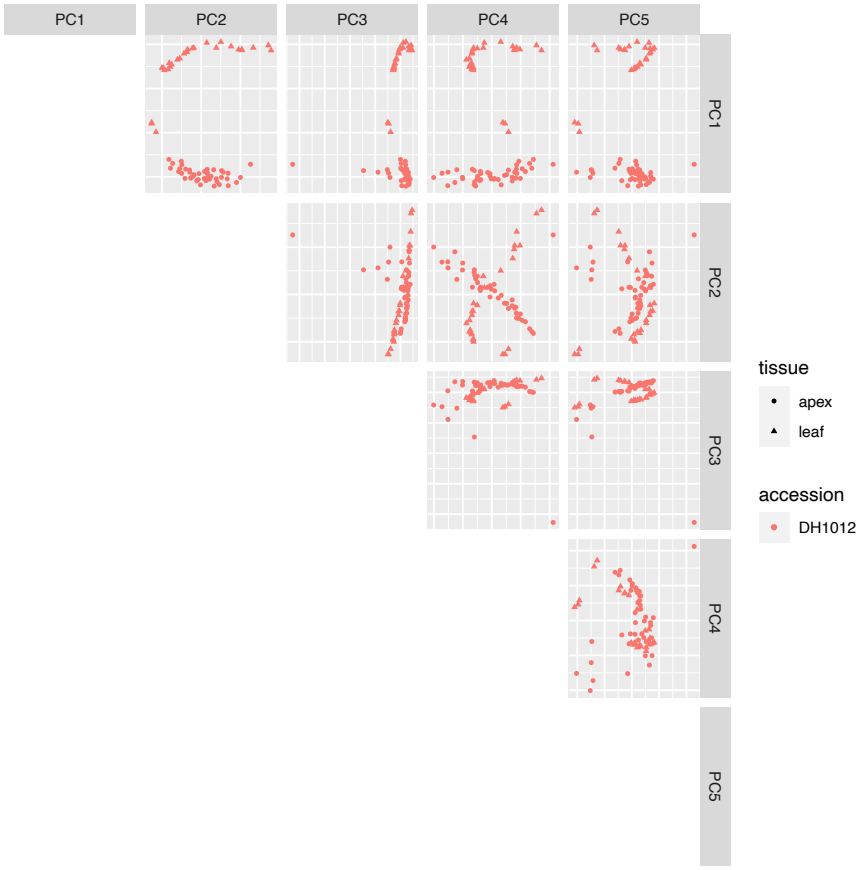


Figure 3.4: Principal Component Analysis (PCA) from PC1-PC5, data from leaf and apex transcripts in *B. oleracea* DH1012.
An outlier was identified in PC1-PC3 and PC1-PC5 as 51 d apex sample, replicate 3.

3.3.5 Differential gene expression analysis at the apex

Differential gene expression (DEG) analysis was carried out using the EdgeR package in R (Robinson, McCarthy and Smyth, 2010). DEG analysis was initially conducted on pairwise comparisons between the earliest timepoint, 14 d, and the remaining timepoints in the time series and this was repeated for both leaf and apex tissues. Two peaks were seen in DEG number across both up and downregulated DEGs at the apex (Fig. 3.5), the first occurred at 33 d, the timepoint just before the apex was seen to exhibit floral primordia under the microscope and the second at 49 d, just before bud emergence. From this, I hypothesised that this first peak could be attributed to the expression of genes involved in shoot apical meristem identity and the second peak the expression of floral organ identity genes. To investigate this expression of key meristem identity genes *AP1, CAL, LFY* and *TFL1* were plotted (Fig. 3.6) in addition to genes known to be involved in floral organ identity, *AG, PI, SEP1, SEP2, SEP3* (Fig. 3.7). The expression was not as predicted and in fact all genes' expression, except for that of *TFL1*, were seen to increase at 39 d, between the two peaks in DEG number. The two

copies of *TFL1* demonstrated different expression patterns to the rest of the genes investigated and this can be attributed to *TFL1*'s role as a floral repressor. *TFL1* expression is known to inhibit the expression of floral promoters *FT* and *AP1*. Here we see *TFL1* expression peak at 31 d, before considerable downregulation in expression from 39 d onward, aligning with the increase in expression of *AP1*.

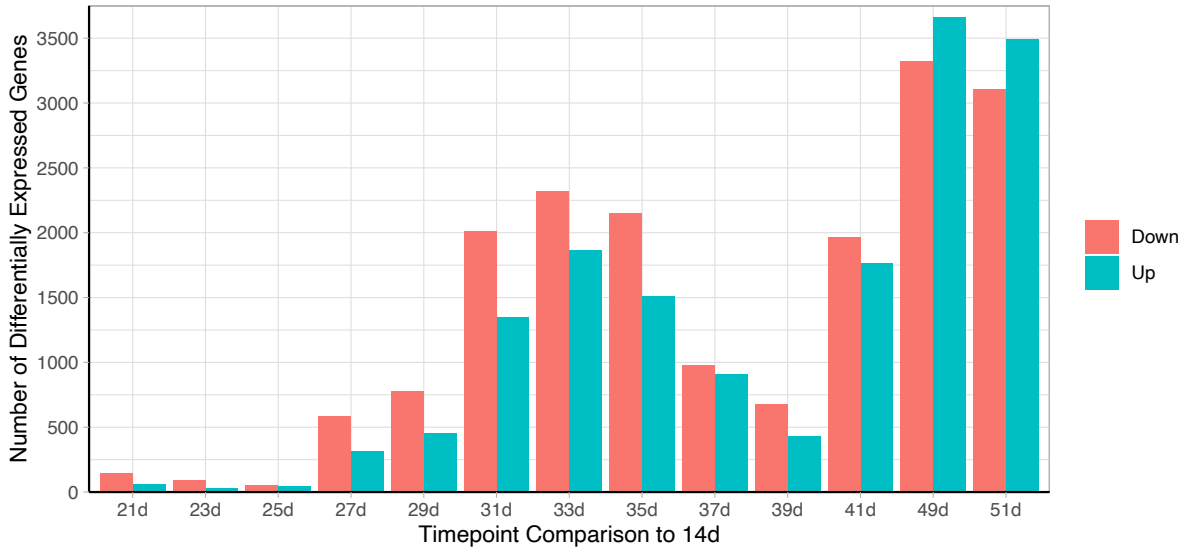


Figure 3.5: Two peaks in the number of both up and down regulated DEGs were seen across time at the apex of DH1012.

DEG number is calculated as the difference in DEGs between each timepoint and the vegetative timepoint 14 d.

As the expression patterns of the genes involved in shoot apical meristem identity and floral organ identity did not align with the double peak in DEG number, further investigation was carried out using GO enrichment analysis. DEGs identified for the 14 d and 35 d comparison and for the 14 d and 51 d comparison were characterised using GO. The upregulated DEGs for the 14 d and 35 d comparison were characterised by 93 GO terms (Table S.5B). The most significantly enriched term was, GO:0048367 shoot system development. Interestingly two other terms that were enriched were GO:0090567 reproductive shoot system development and GO:0009908 flower development terms, both indicating floral development at the apex. 120 GO terms characterised the upregulated DEGs for the 14 d and 51 d comparison (Table S.6B), a higher number of GO terms than for the 14 d and 35 d comparison, reflecting the higher number of DEGs identified. The most significant enrichment term was GO:0000977 RNA Polymerase II regulatory region sequence-specific DNA binding. Again, the flower

development term was enriched; however, a different set of genes led to the enrichment of this term. Between day 14 d and 35 d 87 genes led to the enrichment of GO:0009908, including three copies of *FD,* which are not present in the 14 d and 51 d comparison. For the 14 d 51 d comparison 168 genes led to the enrichment of GO:0009908, including *AP1* and *AP3*. We again also see enrichment of GO:0090567 reproductive shoot system development, in addition to GO:1900618 regulation of shoot system morphogenesis, GO:0048437 floral organ development, GO:0048831regulation of shoot system development, GO:1905392 plant organ morphogenesis and GO:0080154 regulation of fertilisation, all of which point toward a shift to floral organ development, in line with the hypothesis.

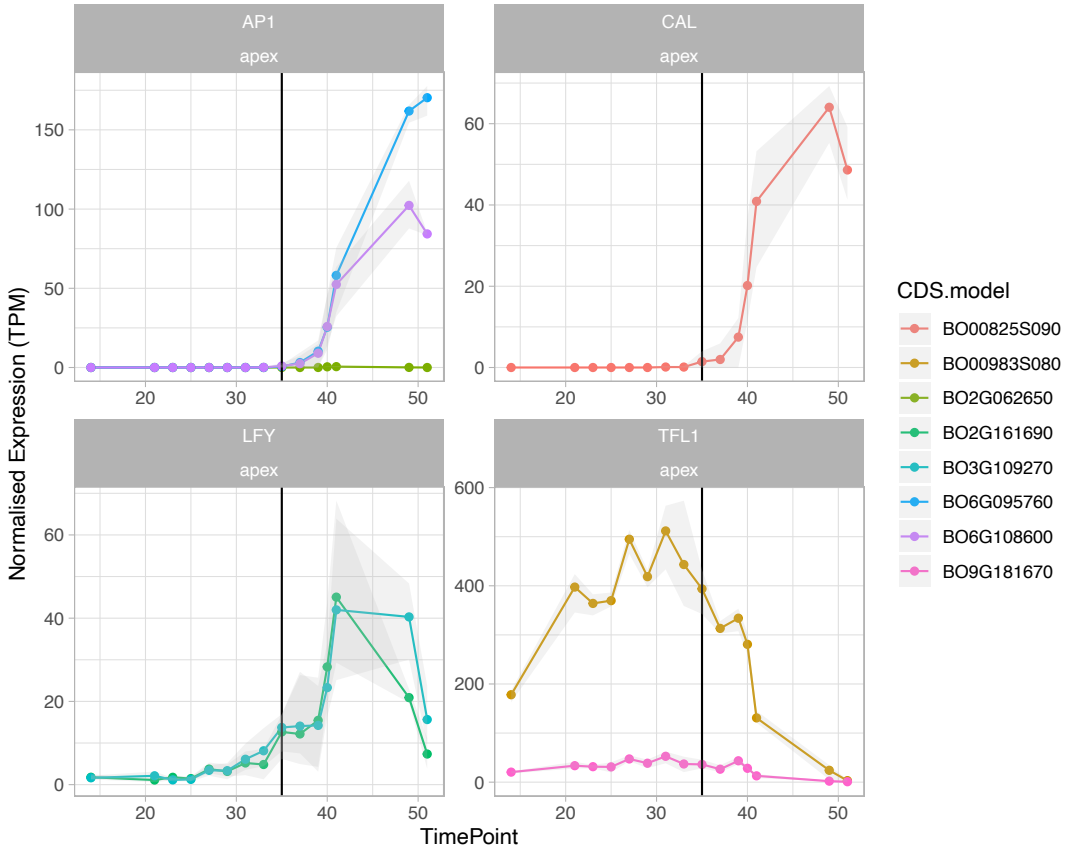


Figure 3.6: Paralogues of *B. oleracea AP1, CAL* and *LFY* are upregulated at 37 d after the meristem has transitioned to an inflorescence meristem.

Gene expression patterns of orthologues of the Arabidopsis shoot apical meristem identity genes, *AP1, CAL, LFY* and *TFL1* across the time series at the apex. Expression measured as TPM; black line indicates day the apex had morphologically transitioned to a floral state.

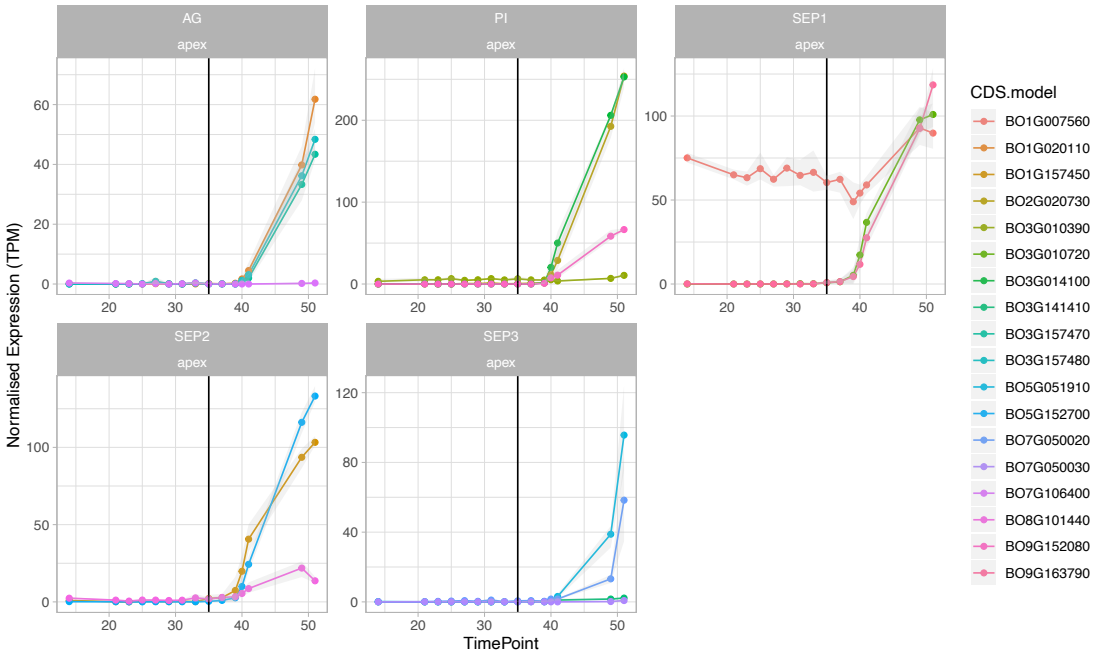


Figure 3.7: Paralogues of *B. oleracea AG, PI* and *SEP1-3* are upregulated at 39 d.
Gene expression patterns of orthologues of the Arabidopsis floral organ identity genes, *AG, PI,* and *SEP1-3* across the time series at the apex. Expression measured as TPM; black line indicates day the apex had morphologically transitioned to a floral state.

To further explore the DEGs in the apex, I compared three key developmental timepoints within the time series (Fig. 3.8), 14 d at which the plants were vegetative, 35 d at which point I had seen the apex was exhibiting floral primordia and 51 d, the final timepoint, at which we saw at least 50 % of the remaining plants had reached developmental stage BBH51. The greatest number of both up and downregulated DEGs were seen between the 14 d and 51 d comparison, 3489 and 3107 respectively. At 51 d the plants are the most developed in the time series and as such it would be expected that many more genes would be expressed at this timepoint, due to their exposure to environmental stresses and because of the activation of the floral transition which requires the expression of numerous genes.

To focus the search for genes involved in the floral transition, I decided to look at how many of these DEGs were orthologues of Arabidopsis Flor-ID genes (Bouché *et al.*, 2015) (Fig. 3.8C, D). This significantly reduced the number of genes in the comparison. Four DEGs were consistently upregulated under all three comparisons (Table 3.1), 14 d and 35 d, 14 d and 51 d, and 35 d and 51 d. The first was BO5G127310, and orthologue of Arabidopsis *SPL5*, AT3G5270. This gene is known to be involved in the ageing pathway in Arabidopsis and therefore it's upregulation over the time series points toward a similar role in DH1012.

Furthermore, orthologues of *AGL6* BO4G024840, *FUL* BO2G161210 and *GA20ox2* BO2G048220 were also seen to be consistently upregulated. The expression of all three is known to promote flowering in Arabidopsis, therefore the upregulation of these orthologues indicates they are performing similar roles in *B. oleracea*.

One gene was consistently downregulated between the three comparisons (Table 3.2), and this was an orthologue of Arabidopsis *AGL19*, BO6G2029290. *AGL19* in Arabidopsis is known to promote flowering, therefore it was not expected to be consistently downregulated across the timeseries. However, it is important to note that DH1012 lacks a vernalisation requirement, and *AGL19* is known to work alongside the vernalisation pathway in Arabidopsis to promote flowering (Schönrock *et al.*, 2006) and therefore may be redundant for the floral transition in DH1012.

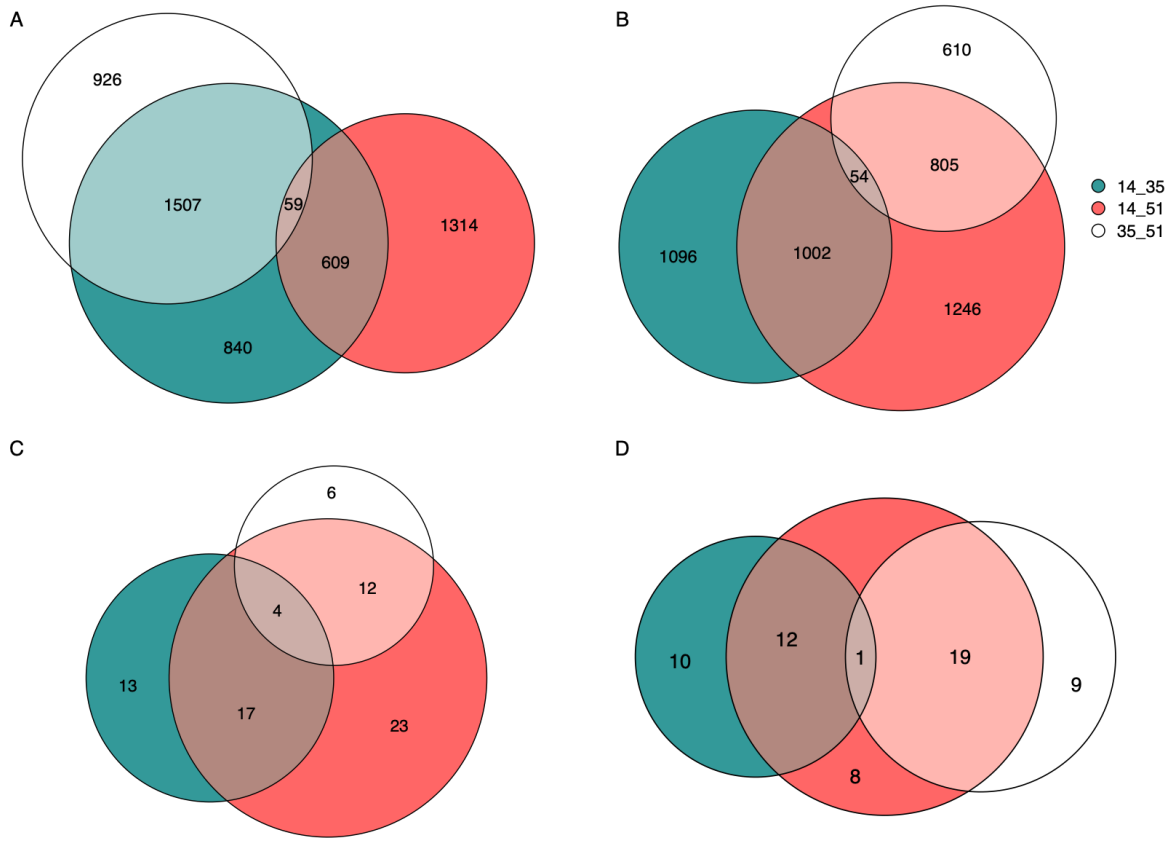


Figure 3.8: Looking at paralogues of the Arabidopsis Flor-ID genes that are differentially expressed greatly reduced the number of candidates.

For the 14 d 35 d, 14 d 51 d and 35 d 51 d comparisons, Venn diagrams to represent A) number of upregulated DEGs B) number of downregulated DEGs C) number of Flor-ID gene paralogues in the upregulated DEG list D) number of Flor-ID gene paralogues in the downregulated DEG list.

Table 3.1: Orthologues of Arabidopsis Flor-ID genes that were seen to be upregulated and differentially expressed for the 14 d 35 d, 14 d 51 d and 35 d 51 d apex comparisons.

CDS Model	Symbol	Arabidopsis Homologue	Gene Description
BO5G127310	SPL5	AT3G15270	Ageing Photoperiodism, light perception and signalling, hormones
BO4G024840	AGL6	AT2G45650	General
BO2G161210	FUL	AT5G60910	Flower development and meristem identity, flowering time integrator
BO2G048220	GA20ox2	AT5G51810	Hormones

Table 3.2: Orthologues of Arabidopsis Flor-ID genes that were seen to be downregulated and differentially expressed for the 14 d 35 d, 14 d 51d and 35 d 51 d apex comparisons.

CDS Model	Symbol	Arabidopsis Homologue	Gene Description
BO6G2029290	AGL19	AT4G22950	Flowering, Transcription, Transcription regulation

3.3.6 Differential gene expression analysis in the leaf

In parallel to the work in the apex, DEG analysis in the leaf was conducted on pairwise comparisons between the earliest timepoint, 14 d, and the remaining timepoints in the time series (Fig. 3.9). On plotting this data, we see a different trend to the apex DEG numbers, here we see a constant increase in both up and downregulated DEGs over time. We took only the first true leaf until it showed signs of senescence and based on current knowledge of the life cycle of leaves, it is to be expected that we would see an increase in DEG number over time. In comparison to the apex data, in the leaf we see significantly more DEGs, both up and downregulated. For the 14 d 35 d apex comparison we see 1508 upregulated DEGs and 2152 downregulated DEGs (Fig. 3.10B). In contrast, in the leaf we see 8296 upregulated DEGs and 8946 downregulated DEGs (Fig. 3.10A). This marked increase in DEG number is to be expected due to the many roles that are performed by leaf tissue. Leaves are the photosynthetic organs of the plant, which requires leaves to be the site of light capture, gaseous exchange and for leaves to have the ability to store sugars and water for the plant. The leaf also has its own

life cycle, with plants producing new leaves over time; the cycle begins with cell division, before moving to cell expansion and eventually senescence, or cell death and all these processes require the expression of different genes.

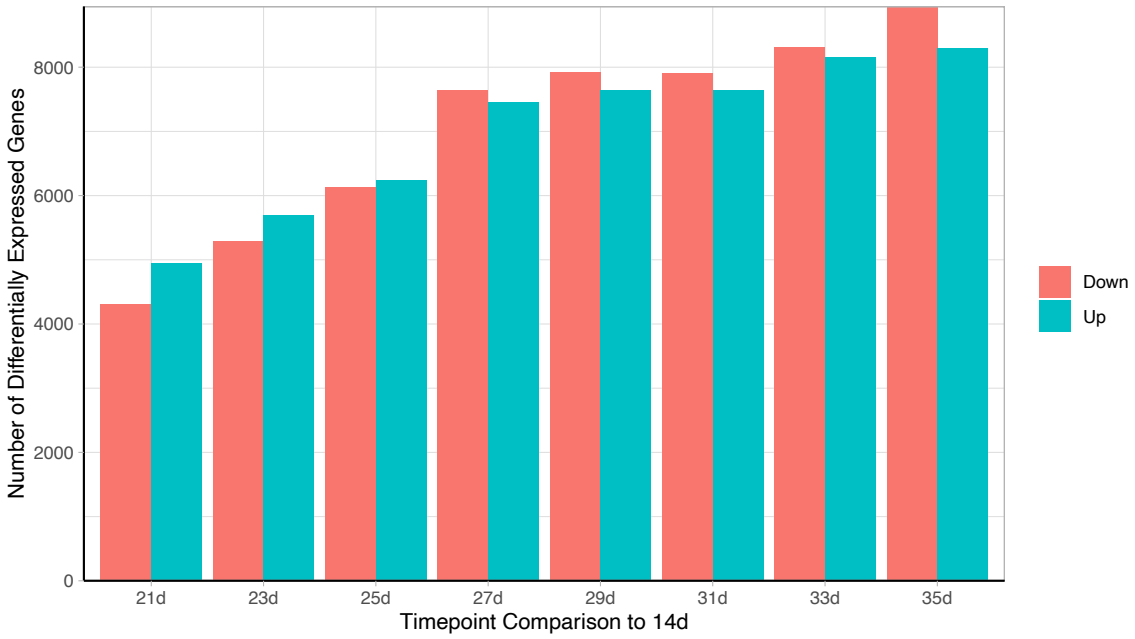


Figure 3.9: DEG number consistently increased until 27d in the leaf tissue.
DEG number is calculated as the difference in DEGs between each timepoint and the vegetative timepoint 14 d.

GO enrichment analysis for the 14 d 21 d and 14 d 35 d comparisons gave an insight into the developmental differences as leaves age. More GO terms were enriched for the 14 d 21 d comparison at 239, compared to 194 for the 14 d and 35 d comparison. For both comparisons the most significantly enriched biological process term was photosynthesis, GO:0015979 and the enrichment of many other terms indicated photosynthesis is the primary task of the leaves under question, for example, GO:0019684 photosynthesis, light reaction, GO:0009522 photosystem I and GO:0009523 photosystem II. However, there were some key differences in the enrichment terms between the two comparisons. For the later comparison we start to see signs of an ageing leaf through the enrichment of terms, such as GO:0007568 ageing and GO:0008219 cell death that we do not see for the 14 d 21 d comparison. No terms were enriched that related to senescence, indicating sampling was stopped prior to senescence becoming an issue with gene expression.

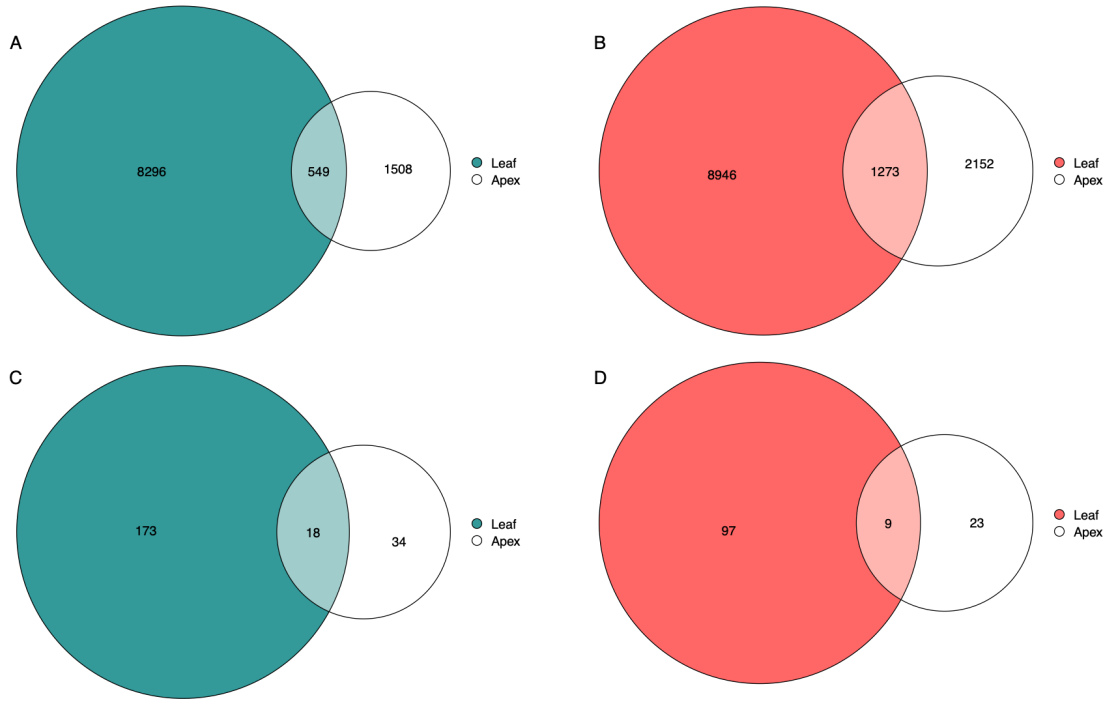


Figure 3.10: Looking at paralogues of the Arabidopsis Flor-ID genes that are differentially expressed greatly reduced the number of gene candidates.

For the 14 d 35 d comparisons in leaf tissue compared to apex tissue, Venn diagrams to represent A) number of upregulated DEGs B) number of downregulated DEGs C) number of Flor-ID gene paralogues in the upregulated DEG list D) number of Flor-ID gene paralogues in the downregulated DEG list.

In Arabidopsis, the leaf has a well-characterised role in the floral transition through the perception of photoperiod, which is largely regulated by *CO* and induces flowering through the expression of the downstream target *FT*. Considering this, the expression patterns of the *FT* copies in DH1012 were investigated as well as its upstream regulator *CO*. A total of four *FT* paralogues were identified in DH1012. In the leaf, one paralogue, BO6G099320, was upregulated over the course of the time series (Fig. 3.11B). The upregulation of this paralogue suggests it is performing a similar role to its Arabidopsis orthologue and promoting flowering. The expression of *CO* paralogue B09G163730 in the leaf tissue, also demonstrated upregulation over time (Fig. 3.11A), starting from the earliest timepoint 14 d, indicating it is performing a similar role to its Arabidopsis orthologue. The CO family contains many genes and most of the *CO*-like gene orthologues found in DH1012 were also demonstrating a similar expression pattern (Fig. S.7B). This was further supported by the expression patterns of downstream targets of Arabidopsis *FT, AP1* and *SOC1* (Fig. 3.11C, D). Upregulation of two of the *AP1* paralogues was observed from 39 d, which is after upregulation of *FT* was seen in

the leaf, as expected with current knowledge from Arabidopsis. Furthermore, two *SOC1* paralogues are also upregulated following the upregulation of the *FT* paralogues. Conversely, the expression patterns of the remaining three *FT* paralogues were very different. BO01129S030 was upregulated over the time series, but its expression was minimal and the remaining two paralogues, BO2G051350 and BO4G061100, were not expressed in the leaf tissue across the time series (Fig. 3.11B).

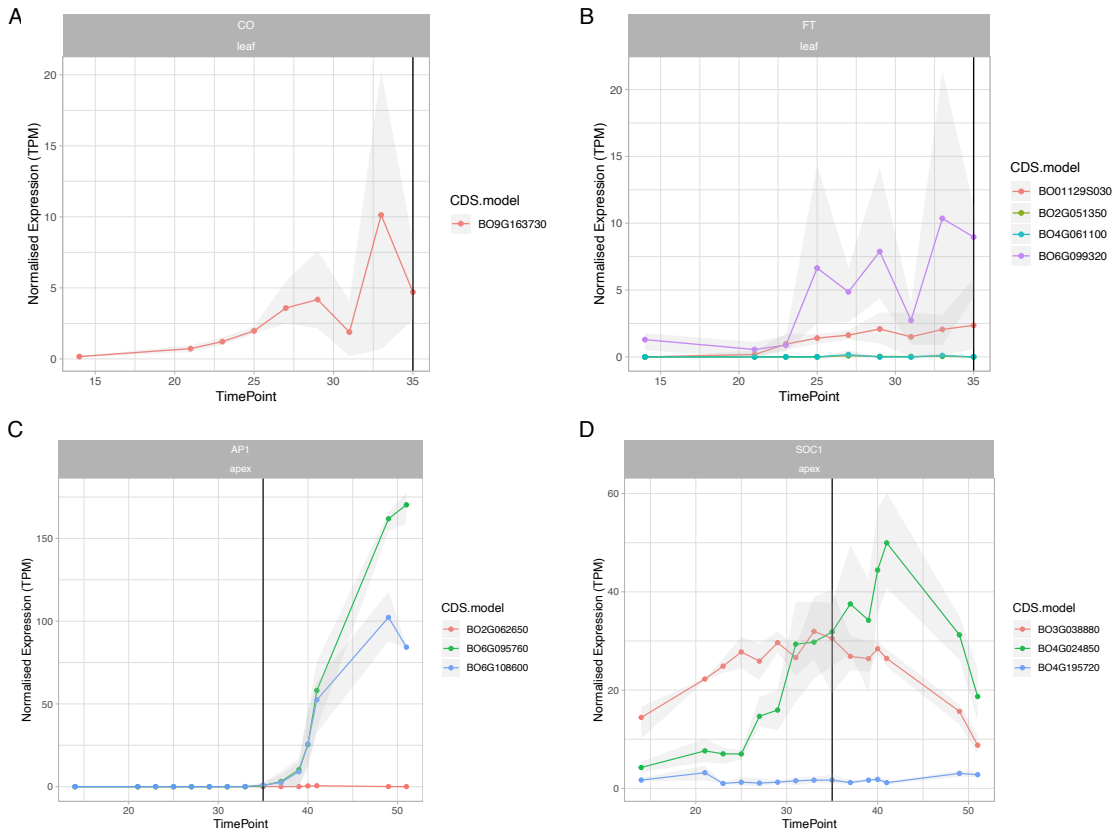


Figure 3.11: Two paralogues of *B. oleracea FT* demonstrate expression patterns that support a conserved role to their Arabidopsis orthologue.

Expression of paralogues of *B. oleracea* A) *CO* B) *FT* C) *AP1* D) *SOC1* over time. Expression measured as TPM, black line indicates the day the apex had morphologically switched an inflorescence meristem, 35 d.

We aligned the genomic sequences for these four *B. oleracea FT* paralogues to the Arabidopsis *FT* genomic sequence (Fig. S.8B) and generated an association tree (Fig. 3.12) to show the evolutionary relationship between the genomic sequences of Arabidopsis *FT* and its *B. oleracea* orthologues. This revealed that BO6G099320 had the highest alignment rate at 60.7%. This added weight to the hypothesis that this paralogue may be performing a similar

role to its Arabidopsis homologue. Furthermore, alignment of the paralogue BO01129S030, which exhibited minimal expression in the leaf over this time series experiment, to the Arabidopsis *FT*, revealed a 417 bp deletion. To explore this further and determine whether this resulted in a non-functional protein, the protein sequences of each of the *B. oleracea FT* paralogues were aligned to Arabidopsis FT (Fig. S.8B). Alignment revealed the deletion of 19 amino acids from the sequence and an overall alignment of ~ 60%. Alignment of the remaining three paralogues all exceeded 81%.

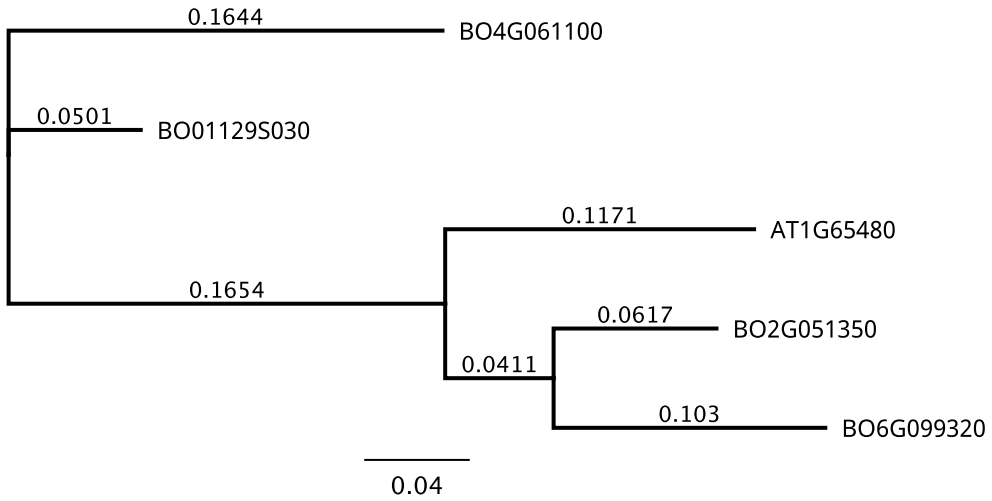


Figure 3.12: Association tree to demonstrate the relationship between the Arabidopsis *FT* and its *B. oleracea* orthologues.
Tree constructed using the Neighbour joining method and genomic sequence data. Branch labels denote the substitutions per site.

3.3.7 Self-organising map-based clustering of expression data

To assess trends in the data, self-organising map (SOM) based clustering was used (Fig. 3.13). SOMs are an unsupervised machine learning technique, in which a high dimensional data set is represented at a lower dimension, here in a two-dimensional format, whilst maintaining the topological structure of the data. In using SOMs to cluster gene expression over time, each cluster represents an expression profile and genes with a similar epression profile will be grouped into that cluster. The training process for SOM clustering means that neighbouring clusters often demonstrate similar expression profiles. Euclidean distances are used to assign neighbours, so neighbouring clusters will have a lower Euclidean distance between each other.

A 3 x 3 SOM was generated using the transcriptomic time series data for the apex tissue (Fig. 3.13A). The SOM is toroidal, such that clusters on the top and bottom rows are adjacent, as are clusters on the left- and right-hand columns. Within this SOM, cluster 6 had the highest number of genes mapping to it, at 1460. The expression profile for cluster 6 begins very low during the vegetative phase and in the final few timepoints, just before bud emergence, we see a significant increase in expression followed by a plateau in expression at the final timepoint as the buds emerge. This suggests there are many genes that are expressed toward the end of the time series, which supports earlier findings on DEG number across the time series. Bud emergence, which occurs at the end of the time series, is a process that requires the coordinated expression of many genes to carry out. Within cluster 6 we see three *FD* paralogues, BO3G156810, BO7G117660 and BO1G006110. *FD* produces a bZIP protein that acts as a positive regulator of flowering through interaction with FT. Within this cluster we also see paralogues of the floral promoters *SOC1* BO4G024850, *AGL24* BO1G039080 and *GA20OX3* BO9G175240. Furthermore, there are paralogues of members of the ageing pathway present within this cluster, *TOE3* BO7G064000 and *SPL4* BO6G029290.

To determine whether the leaf would display expression dynamics similar to the apex, a SOM was also generated for the leaf transcriptome (Fig. 3.13C). Due to the higher number of DEGs identified for the leaf tissue, 17244 genes compared to 8698 for the apex tissue, a 4 x 4 toroidal SOM was produced for this data. Three regions of the leaf SOM had high numbers of genes mapping to them, these were represented by clusters 5, 9 and 13. Cluster 5 exhibits an expression profile that increases over the time series before plateauing at the final timepoints. Interestingly, this cluster included a paralogue of the floral repressor *SVP*. Within Arabidopsis, *SVP* represses *FT* expression by binding to the *FT* promoter (Li *et al.*, 2008), so it is interesting that this paralogue is demonstrating upregulation over the time series, as we would expect it to be downregulated as a floral repressor indicating that this copy could be performing a different role within *B. oleracea*.

Cluster 9 exhibits an expression pattern that rapidly declines at the beginning of the time series, during the early vegetative phase, and then a slower decline for the remaining timepoints. This indicates that these genes may be important in the early development of the plant, prior to 14 d. Cluster 13 is the largest cluster with 934 genes mapping to it. This cluster exhibits an expression pattern not dissimilar to an exponential decay curve with a rapid initial decline in expression over the vegetative phase, before levelling off toward the end of the time series during the floral phase of development. This cluster includes a paralogue of the Arabidopsis gene *AGL19* BO7G108370. A different *AGL19* paralogue was found in earlier DEG analysis as being consistently downregulated across the time series, and it appears that

this paralogue is also downregulated over the time series, despite its Arabidopsis counterpart promoting flowering (Schönrock *et al.*, 2006). This further suggests that due to the lack of vernalisation requirement in DH1012, *AGL19* paralogues may be performing a different function in DH1012 or be redundant.

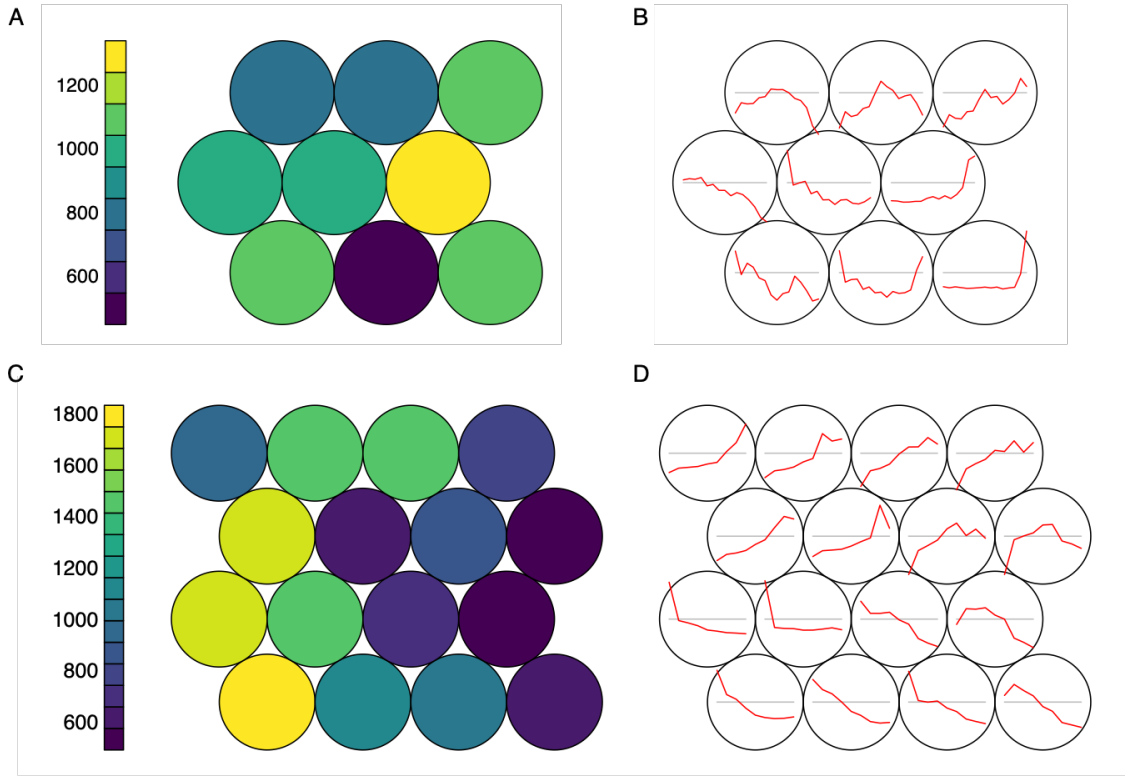


Figure 3.13: SOM clustering revealed similar expression dynamics between leaf and apex tissues.

3 x 3 toroidal SOM for apex A) number of genes per cluster B) normalised expression profile represented by that cluster. 4 x 4 toroidal SOM for leaf C) number of genes per cluster B) normalised expression profile represented by that cluster.

3.4 Discussion

Developing a predictive understanding of the floral transition in *B. oleracea* is a key goal for growers and breeders to aid future breeding strategies. From studies in Arabidopsis, it is known that the leaf and apex tissues are responsible for different aspects of the floral transition (Srikanth and Schmid, 2011; Andrés and Coupland, 2012). To investigate flowering in *B. oleracea* DH1012, we therefore collected transcriptome time series in both leaf and apex tissues.

We first used the transcriptome time series data to detect and quantify the differences in gene expression between the leaf and apex tissues. Analysis of DEG number in the apex over time revealed two peaks for both up and downregulated genes. The first peak coincided with the morphological transition of the apex from a vegetative meristem to an inflorescence meristem, whilst the second, larger DEG peak coincided with bud emergence within the plants. An investigation was conducted into orthologues of known Arabidopsis meristem identity genes and floral organ identity genes to see if their expression patterns correlated with these findings. Based on knowledge from Arabidopsis we expected that the meristem identity genes expression would correlate to the first DEG peak, whilst expression of the floral organ identity genes would coincide with the second peak. This was not the case and all of the investigated genes, with the exception of floral repressor *TFL1*, were upregulated at 39 d between the two DEG peaks. As the expression of the investigated meristem identity genes was upregulated after the meristem had morphologically transitioned into a floral state, this suggests that other genes could be responsible for the transition of the meristem in DH1012. The increase in the expression of these meristem identity genes coincides with the second DEG peak observed and could suggest that in DH1012, these genes are involved in the later stages of the floral transition, for example floral organ development.

To understand which biological processes were coinciding with the peaks in DEG numbers, GO enrichment analysis was conducted. This revealed that for the DEGs under the first and second DEG peak, the GO:0009908 flower development term was enriched, but the DEGs underlying these terms were different. All three *FD* paralogues were present in the first peak, indicating that in DH1012, *FD* could play a role in the initial morphological switch from a vegetative to an inflorescence meristem. *FD* was not found to be differentially expressed at the later second peak in DEG number. This makes *FD* a strong candidate for involvement in meristem identity within DH1012, Within Arabidopsis it has been shown that *FD* plays a role in maintaining inflorescence meristem identity in an age-dependant manner, through its interaction with *FT* (Gorham *et al.*, 2018).

Interestingly, all orthologues of Arabidopsis meristem identity and floral organ identity genes saw a marked increase/decrease in expression at day 39. From this it could be speculated that this is a critical timepoint in the development of DH1012. Such a shift in gene expression states indicates that major change is occurring within the plant, perhaps a reprogramming is occurring at the genetic level at this time. If this hypothesis is correct, this then poses the question, which genes are causing this morphological change at the meristem, as 39 d is after the switch to an inflorescence meristem has occurred. This would mean that the gene

regulatory network controlling the floral transition would look vastly different between DH102 and Arabidopsis.

A similar pattern of DEG numbers was observed in a transcriptome time series experiment within Arabidopsis (Klepikova *et al.*, 2015). In this study of Arabidopsis, apex tissue was collected from 7 d to 16 d. The first peak in DEG number was 11 d post-germination and the second peak was at 16 d, the final time point in the series. Using clustering and GO enrichment, the first peak was attributed to the transition to the floral phase, or more specifically the development of an inflorescence meristem, and the second was attributed to flower development, in alignment to findings here in DH1012.

Within the leaf a different pattern in DEG number was seen to the apex. DEG numbers in the leaf are much higher overall and the numbers of DEGs increase across time. This was to be expected due to the numerous roles leaf tissues fulfil (Lauri and Normand, 2017). Aside from their critical roles in photosynthesis, plant defence and in flowering time, all processes that require the coordinated expression of many genes, the leaf has its own life cycle and is ageing over time. GO enrichment analysis revealed the onset of genes involved in ageing and cell death by the final sample, through the enrichment of GO:0007568 ageing and GO:0008219 cell death. Despite this, the enrichment of terms describing photosynthesis remained consistent across the time series. This was expected, as photosynthesis is the primary function of the leaf.

To explore the role of the leaf in DH1012 during the floral transition, the expression patterns of paralogues of Arabidopsis *FT* were investigated, alongside the upstream regulator *CO* and downstream targets of *FT*: *AP1* and *SOC1*. Understanding the role of *FT* in *B. oleracea* is particularly interesting in light of recent work on *B. rapa* that demonstrated that the rapid cycling variety Sarisha-14 appears to bypass the photoperiod pathway and *FT,* in favour of the ageing pathway to initiate the floral transition (Calderwood, Hepworth, *et al.*, 2021). DH1012 is also a rapid cycling variety, requiring no vernalisation to initiate flowering, so I hypothesised that it too would bypass *FT* in favour of the ageing pathway to trigger flowering. In Arabidopsis, *FT* expression is promoted by the upregulation of the floral integrator gene *CO.* In DH1012, only one *CO* paralogue was identified, and its expression pattern demonstrates upregulation consistent with the subsequent upregulation of *FT,* suggesting its function is conserved between Arabidopsis and *B. oleracea.*

In Arabidopsis, *FT* plays a key role in the photoperiod pathway as a promoter of flowering (Kobayashi *et al.*, 1999) and within DH1012 four paralogues of *FT* have been identified. The

expression of *FT* paralogue BO6G099320 is similar to Arabidopsis *FT*, indicating that its role may be conserved. Furthermore, sequence analysis of the *B. oleracea FT* paralogues alongside Arabidopsis *FT*, revealed sequence similarity across the paralogues. Alignment of protein sequences revealed > 80% alignment for BO2G051350, BO4G061100 and BO6G099320, although an alignment rate of ~ 60 % was seen for the BO01129S030 copy due to a deletion. The similarities in expression profiles combined with the sequencing results suggest a conserved function between BO6G099320 and the Arabidopsis *FT*, but functional analysis of this *FT* paralogue, through complementation experiments in Arabidopsis *ft*-mutants, would confirm whether this hypothesis is true.

To study the differences in gene expression dynamics between the leaf and apex tissue, SOMs were generated to cluster the expression profiles over time. Three main expression responses were observed in both the apex and leaf tissues, the first was a rapid decline in expression at the first few timepoints followed by low expression for the remainder of the time series, indicating the expression of genes involved in the early vegetative stage which are not needed later in development. In contrast, the second was low expression followed a large increase in expression toward the final few timepoints, in which bud emergence occurred, suggesting the onset of expression of genes involved in floral organ identity. Finally, the third pattern was a marked increase at the middle of the time series, the time at which the apex transition to a vegetative state, indicating the onset of expression of genes involved in meristem identity. These findings align with results from the initial DEG analysis of the apex transcriptomes in which there are peaks in DEG expression level around the time of floral transition in the apex and at bud emergence.

3.5 Conclusion

Here we provide a comprehensive study of gene expression during the floral transition in *B. oleracea* DH1012. A marked increase in DEG number was seen at two key developmental stages in plant development, the transition of the vegetative meristem to an inflorescence and at the point where buds began emerging in the remaining plants. GO enrichment analysis revealed that these increases in DEG numbers coincide with the expression of genes involved in meristem identity and floral organ identity respectively. Furthermore, we have determined that unlike in *B. rapa* variety Sarisha-14, at least one *FT* paralogue is important for the initiation of flowering in *B. oleracea* DH1012, indicating that the photoperiod pathway is important despite the lack of need for the vernalisation pathway in this rapid cycling line.

Chapter 4

Using comparative transcriptomics to gain a better understanding of the floral transition in *B. oleracea*

4.1 Introduction

4.1.1 The transfer of knowledge from model to crop will provide a valuable resource for plant breeding

B. oleracea and Arabidopsis are both members of the *Brassicaceae* family that diverged from a common ancestor ~ 43.2 Mya. The conservation of large chromosomal chunks is evident within these two species, alongside a persistence of shared chromosomal rearrangements (Ziolkowski *et al.*, 2006; Beilstein *et al.*, 2010). It is therefore likely that orthologous genes will perform similar functions in both species. Indeed, this has been demonstrated with an orthologue of *FRI, BoFRIa,* which through complementation studies has been shown to be functional in Arabidopsis (Irwin *et al.*, 2012). Furthermore, orthologous copies of *FLC* and *FT* have been identified as strong candidates for flowering time regulation in *B. oleracea* (Pires *et al.*, 2004; Lin *et al.*, 2005; Razi *et al.*, 2008; Ridge *et al.*, 2015; Irwin *et al.*, 2016). However, a large gap in current knowledge remains regarding the differences in expression dynamics for the floral transition genes between the two species and the regulatory interactions controlling flowering. Due to the wealth of knowledge on flowering in the relative of *B. oleracea,* Arabidopsis, the successful transfer of knowledge between the species would be an important step for future studies.

A major limitation in the field of evolution of organism morphology is the ability to make meaningful comparisons between different species. One study, used the model plant Arabidopsis to compare to *B. rapa* R-o-18 (Calderwood, Hepworth, *et al.*, 2021). Much like *B. oleracea, B. rapa* is a close relative of Arabidopsis and as a species contains many important crops, *B. oleracea* and *B. rapa* are thought to have diverged only ~ 4 Mya (Rana *et al.*, 2004). Through the development of a method to translate the Arabidopsis gene expression profiles onto those of *B. rapa*, this study enabled the investigation into many orthologues of

Arabidopsis flowering time genes to compare their expression profiles to those in *B. rapa*. It was found that different genes required different translations for the expression profiles between the two species to be aligned. This suggested that the increased developmental time seen within the *B. rapa* cannot be explained by a blanket transformation of all genes. This observation points towards differences in the gene regulatory networks for flowering between *B. rapa* and Arabidopsis, or rather that there is no common developmental time between Arabidopsis and *B. rapa*.

Numerous functional orthologues of Arabidopsis flowering time genes have been identified in crop species with conserved functions (Blümel, Dally and Jung, 2015). Understanding the floral transition is critical to agriculture, and our extensive knowledge in the model species Arabidopsis could be a valuable resource for plant breeding if we are able to successfully transfer it into crop species (Jung and Müller, 2009) such as *B. oleracea*.

4.1.2 The presence of multiple gene copies drives neo- and subfunctionalisation

B. oleracea is a diploid species, however a whole genome triplication event that took place in its evolutionary history means that the presence of multiple gene copies, is a common occurrence within its genome. Polyploidy is a prevalent phenomenon in the plant world, and eukaryotic evolution as a whole, leading to increases in complexity and adaptive radiation and speciation. Despite its prevalence, more often than not, gene copies are often analysed together rather than as individual entities. This assumes that the gene copies are performing similar functions or at least have similar expression profiles and does not take into account the possibility that evolutionary pressure on these genes may have led to neo- or subfunctionalisation.

A study carried out in *B. napus* aimed to investigate the retention of Arabidopsis flowering time homologues in *B. napus* and to look at their regulation (Jones *et al.*, 2018). Not only did this study reveal that flowering time genes were preferentially retained within *B. napus,* but it also went on to explore the expression dynamics of some specific examples of flowering time genes that exhibit multiple copies within *B. napus*. There are four copies of *BnaLFY*, and all were initially lowly expressed in the apex, before increasing following vernalisation, analogous to Arabidopsis *LFY*. The similar expression dynamics of these homologues was consistent with the gene balance hypothesis (Papp, Pál and Hurst, 2003; Birchler and Veitia, 2010). This hypothesis states that there is preferential retention of dosage sensitive genes

following a whole genome duplication event, which we know has happened in the evolutionary history of *B. napus*. In contrast, the expression dynamics of *BnaTFL1* homologues, of which there are four, were all distinct from one another, suggesting there had been some neo- or subfunctionalisation between the copies.

4.1.3 Gene Regulatory Network inference

Gene regulatory networks (GRNs) describe the interactions between a set of genes and can be used to provide the molecular insights into biological processes (Penfold and Wild, 2011; En Chai *et al.*, 2014; Chan, Stumpf and Babtie, 2017; Mercatelli *et al.*, 2020). GRNs contain specific genes that code for transcription factors, which target DNA and bind upstream of target genes in regulatory regions, to control their transcription. The interactions between genes and gene products are critical in many molecular processes. Complex GRNs underly many developmental transitions within *B. oleracea* and these networks perceive and process environmental and endogenous signals. Using transcriptome time series data, the inference of GRN topology has proven useful for the unravelling of numerous complex biological processes (Marbach *et al.*, 2012; Jaeger *et al.*, 2013; Penfold and Buchanan-Wollaston, 2014). However, there is no standard method for the generation of GRNs, and many computational approaches have been developed to infer them (Penfold and Wild, 2011; En Chai *et al.*, 2014).

Many approaches reverse engineer GRNs from transcriptome data. High-throughput sequencing techniques allow for the expression profiles of genes of interest to be experimentally determined. Methods for the reverse engineering of GRNs include Bayesian network methods, Boolean network methods, differential equation methods, amongst others (Marbach *et al.*, 2012; En Chai *et al.*, 2014; Saint-Antoine and Singh, 2020; Zhao *et al.*, 2021). Bayesian, Boolean, differential equation, and neural network methods work by constructing a model to describe the relationship among the genes of interest and then fitting the gene expression profiles of these genes to the model to generate a GRN. These methods are highly scalable and flexible and are able to choose different models according to the data type (Zhao *et al.*, 2021).

Many methods for inferring connections between genes use correlation to group genes based on similar expression patterns, the higher the correlation between genes, the higher the probability of interaction between them. Small amounts of data can be used to produce large networks with this method and the method is of low computational complexity (Song, Langfelder and Horvath, 2012; Zhao *et al.*, 2021). Mutual information, conditional mutual information and Pearson correlation coefficient are commonly used metrics. The final group

of methods use machine learning; they use data structures paired with machine learning algorithms to fit gene expression data. These methods allow for directionality, so the inferred GRN is often a directed network and is easy to interpret (Li, Wu and Ngom, 2018; Zhao *et al.*, 2021). Random forest and boosting are commonly used machine learning algorithms for GRN production.

Numerous GRN have been constructed to describe the floral transition in Arabidopsis. One such network describes just the core genes involved in the transition and has been used to predict the flowering behaviour of different genotypes (Jaeger *et al.*, 2013). The network itself was constructed using an ordinary differential equation framework but was influenced by biological observations, which allowed the model to be trained to the data available and meant it could be related back to biological entities. The differences in the GRNs behind flowering in Arabidopsis and *B. rapa* have recently begun to be explored (Calderwood, Hepworth, *et al.*, 2021). Using Causal Structural inference (CSI) (Penfold *et al.*, 2015), which uses a Gaussian process model to build a GRN from gene expression profiles, GRNs were generated to describe the interactions between the key floral genes *FLC, FUL, SOC1* and *SVP* in Arabidopsis and *B. rapa* (Calderwood, Hepworth, *et al.*, 2021).

4.1.4 Hypotheses and Aims

To investigate the genetic regulation of the floral transition of *B. oleracea* in more detail, this chapter employs the curve registration technique developed by Calderwood *et al.* 2021 (Calderwood, Hepworth, *et al.*, 2021) to the transcriptome time series data described in Chapter 3. In *B. rapa,* this method revealed that gene expression dynamics are similar to those of Arabidopsis for many genes, but the timing and magnitude of the expression differs. I hypothesise that this will also be the case for many *B. oleracea* genes and therefore the core gene regulatory network for flowering will be similar to the Arabidopsis network.

Furthermore in *B. rapa*, it was found that genes with similar registration parameters were involved in similar biological processes. Due to the close relationship between *B. oleracea* and *B. rapa*, I hypothesise that this will also be true for *B. oleracea* and will test this using gene ontology enrichment analysis.

It has been documented that the presence of multiple gene copies can lead to neo- and subfunctionalisation between gene paralogues (Birchler and Veitia, 2010). Using the curve registration method, it is possible to clearly compare expression profiles between species and I predict that using this method differences in the expression dynamics of paralogues many

flowering time genes will be seen and this will aid our understanding of the roles of these genes in *B. oleracea.* Specifically, due to previous work carried out in *B. napus* (Jones *et al.*, 2018) and the differences in expression between paralogues, I hypothesise *B. oleracea* orthologues of *TFL1* will demonstrate expression differences.

4.2 Methods

4.2.1 Gene selection for gene registration

Orthologues of Arabidopsis genes in DH1012 were identified (3.2.2). Using this information, pairs of orthologous genes with reproducible, variable expression across the time series were selected to identify differences in gene expression over time between Arabidopsis and *B. oleracea*. For each pair, the variance in expression as counts per million (CPM) for each gene was estimated over time, and if the variance explained was greater than 0.7 in both organisms, the gene was selected. The reason for this was to select genes for which, relative to variation between replicates within each timepoint, the mean expression changes by a large amount between timepoints. This resulted in a comparison of 1988 DH1012 genes to 1187 Arabidopsis genes.

Using these gene sets, the gene expression distance between samples was calculated. To do this, the mean squared difference in gene expression was calculated for orthologous pairs between Arabidopsis and DH1012. This was then scaled per gene to control for the differences in magnitude, by subtracting the mean expression over the time series and dividing by the standard deviation.

4.2.2 Registering gene expression profiles across time

To register gene expression profiles from Arabidopsis to DH1012, the gene registration technique developed by Calderwood *et al.* 2021 (Calderwood, Hepworth, *et al.*, 2021) was used. The gene expression profiles of Arabidopsis were shifted and stretched, and the least squares criterion employed to determine the optimal shift and stretch factors. Using the mean and standard deviation of the overlapping registered points between Arabidopsis and DH1012, the gene expression levels were centred and scaled. Due to the longer developmental time of DH1012 in comparison to the *B. rapa* lines used in the original paper, stretch factors of 1.5 x to 4 x, with an increment of 0.5, in addition to shift factors of -4 and +4 days, with an increment

of 1 day, were tested. For each combination of shift and stretch values at each timepoint for both species, gene expression was linearly imputed between the mean observed values and a score was given using the mean squared difference between the DH1012 timepoint and the imputed Arabidopsis expression value for the overlapping timepoints. Consequently, the best shift and stretch values for a gene would result in a minimal score and would be carried forward to compare to a model without any registration.

To compare between the registered and non-registered models, Bayesian model selection was used. Cubic spline models, using 6 parameters were fit to the overlapping timepoints between Arabidopsis and DH1012 that were identified through the registration process. Bayesian Information Criterion (BIC) (G, 1978; Neath and Cavanaugh, 2012) was used to compare the models for each gene and determine those that had successfully registered.

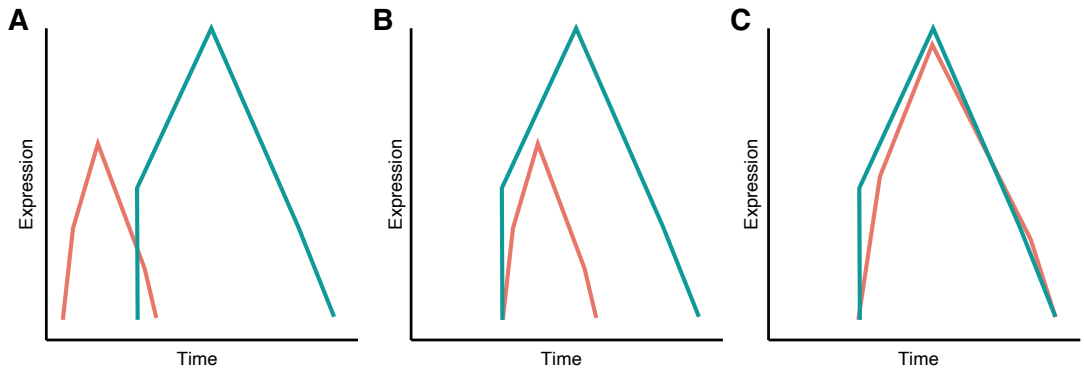


Figure 4.1: Schematic to represent the gene registration process.

A) Gene expression profiles for a homologous gene pair is selected. B) The gene expression profile from the selected species is shifted and C) stretched until it is aligned to the expression profile of the second species. An optimal shift and stretch value is calculated for each homologous pair using Bayesian model selection.

4.2.3 Network Inference using the causal structure inference algorithm

To generate the networks, the likelihood of regulatory links between genes was inferred using CSI v1.0 (Penfold and Wild, 2011), which was run through Cyverse (Merchant *et al.*, 2016). TPM gene expression values of selected genes were used as input and the default parameters were used (parental set depth = 2, gaussian process prior = 10;0.1, weight truncation = 1.0E-5, data normalisation = standardise, zero mean, unit variance, weight sampling = FALSE). The output marginal and MAP files were converted to Graph Modelling Language (GML) format using hCSI_MarginalThreshold v1.0 in Cyverse with a probability threshold of 0.01.

The GML files were used for directed network analysis in Cytoscape v3.8.2 (Shannon *et al.*, 2003).

4.3 Results

4.3.1 Gene expression over time appears dissimilar between Arabidopsis and DH1012

SEM imaging demonstrated that the morphological development at the apex between dh1012 and Arabidopsis Col-0 is very similar but plays out over different time scales (Fig. 4.2). Gene expression over the floral transition was compared between Arabidopsis and DH1012 apex tissue (Fig. 4.3A). Due to the similarities in their morphological development, it was predicted their transcriptomes might display similar gene expression profiles to reflect this. Through comparison it was evident that samples taken at similar times within a species have more similar gene expression profiles than samples at dissimilar times. There are a few exceptions to this general trend, in DH1012 the expression states of the final two timepoints are similar to one another, but very different to the remaining timepoints in DH1012, which may be due to the eight-day difference between the nearest sample within the time series. Furthermore, we see that 33 d is less similar than its surrounding timepoints, this could be due to the fact that it is the timepoint before the floral transition was seen at the apex in DH1012 (3.3.2) and also a peak in DEG number was demonstrated at this timepoint (3.3.4), suggesting it may represent a short-lived gene expression state. Overall, these results suggest the data was sufficient to detect developmental changes through gene expression.

In contrast, no similarity is seen between gene expression states at similar times between Arabidopsis and DH1012. This difference in gene expression states can be partially accounted for by differences in magnitude of expression between the two species. On scaling, more similarity in gene expression at similar timepoints between species was seen, but it was still markedly less than within a species. Additionally, this stark difference with the last two DH1012 timepoints is still visible and likely can be attributed to the larger time difference between these two samples and the rest of the time series. Consequently, it appears there are no similar gene expression states between Arabidopsis and DH1012 across the floral transition, despite the close evolutionary relationship between the two species.

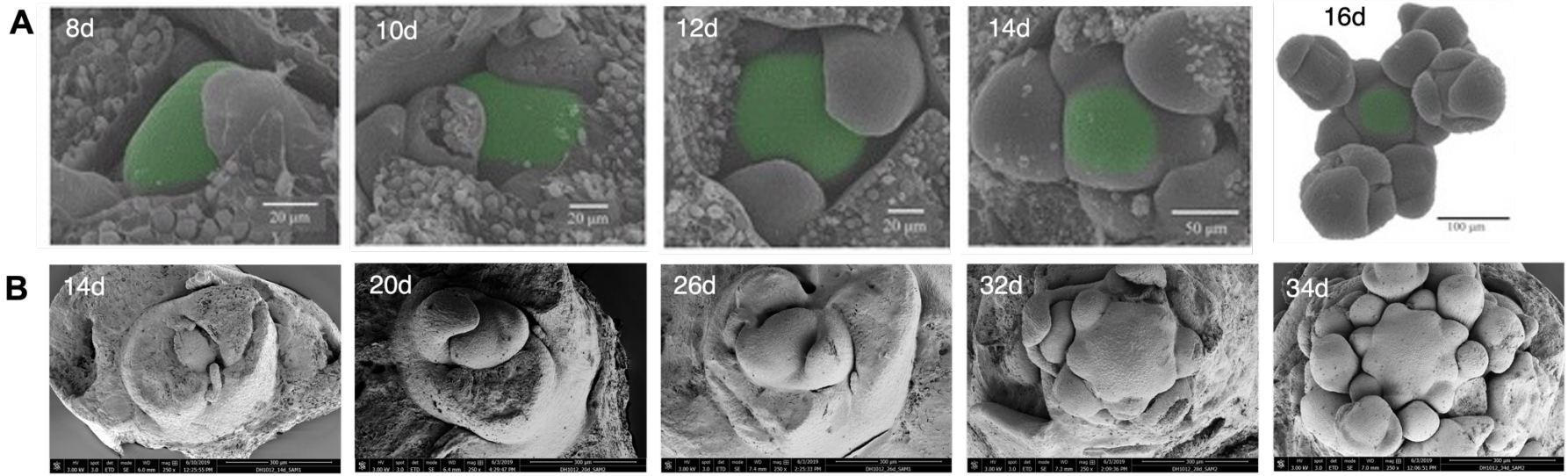


Figure 4.2:The morphological development between Arabidopsis Col-0 and *B. oleracea* DH1012 is similar, but stretched over a longer time period. Scanning electron microscope images showing the developmental stages of the A) Arabidopsis Col-0 and B) the *B. oleracea* DH1012 shoot apical meristem. Col-0 SEM images are reproduced from Klepikova *et al.* 2015.

4.3.2 Registration indicates that gene expression between Arabidopsis and DH1012 is similar but differently synchronised

To determine whether desynchronisation of the gene expression profiles could explain the differences in transcriptomic gene expression between Arabidopsis and DH1012, curve registration was employed. Of the 1988 *B. oleracea* genes investigated using curve registration, the BIC for 1185 favoured a model that considers gene expression in DH1012 and Arabidopsis to be the same after gene registration, over a model in which they are considered to have different gene expression profiles (Fig. 4.3C). This equated to ~ 60 % of genes registering, which was comparative to the 62 % of genes registered in the original *B. rapa* comparison (Calderwood, Hepworth, *et al.*, 2021). This supports the hypothesis that the differences in gene expression between DH1012 and Arabidopsis can be explained by differences in timing rather than in differences in the expression profiles. The heatmap further demonstrates this point. It is now possible to see an increase in the distance in gene expression space between dissimilar timepoints and a reduction in the distance between nearby timepoints. The trend points toward a uniform progression from early to late gene expression states in both Arabidopsis and DH1012, a trend that could not have been detected using the unregistered expression profiles.

Interestingly, the registered genes had different optimal registration parameters, suggesting that there is no single progression through transcriptomic states that is common to both species. This indicates that it is not possible to map Arabidopsis and DH1012 to a common developmental time, as no equivalent developmental stage has been identified at the transcriptomics level between the two. However, through GO enrichment analysis it was demonstrated that genes with similar optimal registration parameters were largely enriched in similar GO terms. Of the 43 registration parameters used, 23 were enriched in less than five GO terms. This indicates that these genes, as well as exhibiting synchronised expression differences between the Arabidopsis and DH1012 time series, may also be involved in similar biological processes or have similar functions.

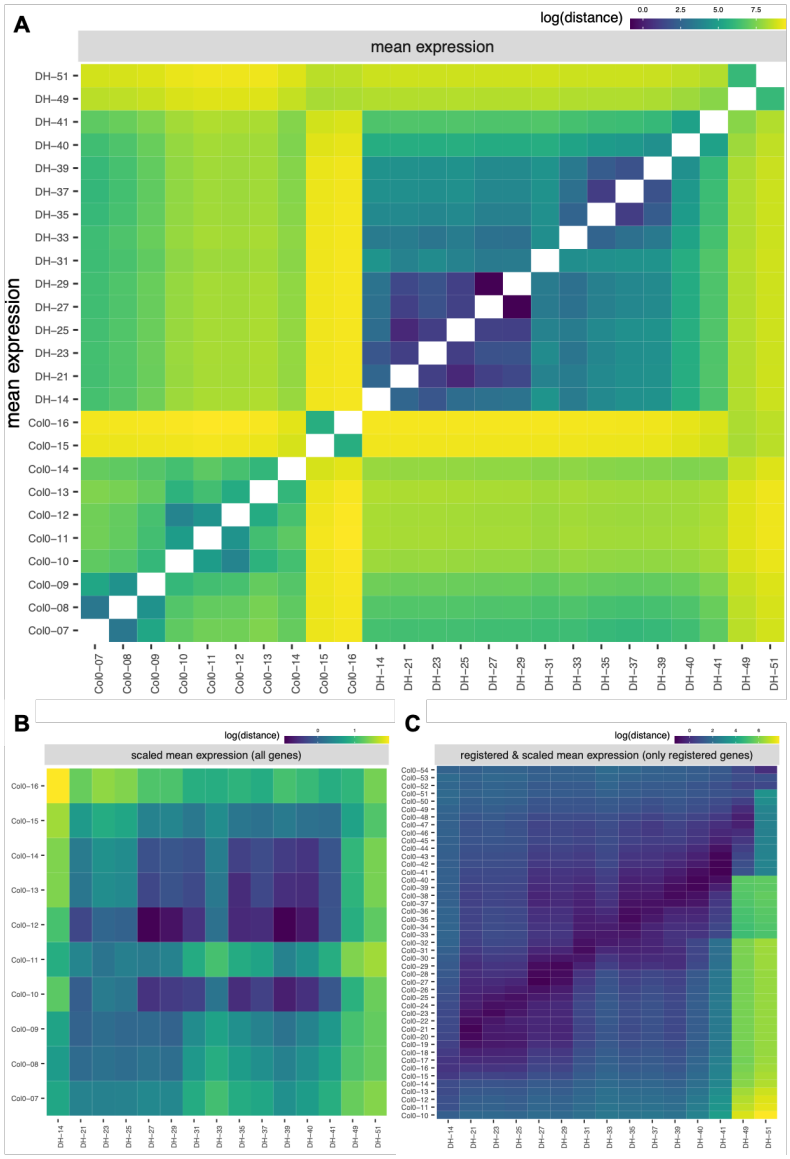


Figure 4.3: Curve registration enables the resolution of differences in gene expression states at the shoot apical meristem between Arabidopsis and *B. oleracea*.

Heatmaps to show the gene expression distance of samples taken from Arabidopsis and DH1012 over time since germination. The average squared difference in expression between homologous gene pairs is used as the gene expression distance. A) Between species gene expression is not similar over time, no obvious structure is seen. Comparison within species (upper-right and lower-left quadrants) reveal similar timepoints have similar expression. This indicates that despite their similar morphological states, there is not similar gene expression between Arabidopsis and DH1012. B) Scaling the gene expression to control of magnitude of expression enables a pattern to begin to emerge, there is more correlation between species samples at similar timepoints. C) Curve registration demonstrates that differences in gene expression are likely to stem from desynchronisation rather than different expression patterns in most cases.

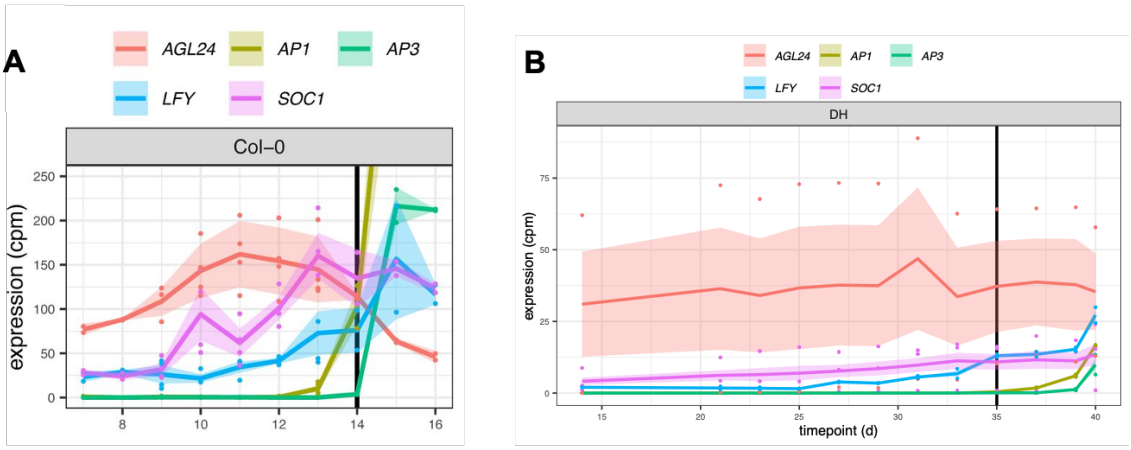
4.3.3 Investigating the expression of key floral genes in Arabidopsis and *B. oleracea*

Mirroring the work previously undertaken in *B. rapa* (Calderwood, Hepworth, *et al.*, 2021), the gene expression profiles of five key genes between Arabidopsis and DH1012 were investigated. (Fig. 4.4) The genes selected were *AGL24, AP1, AP3, LFY* and *SOC1,* as the expression patterns of these genes are diagnostic for different developmental stages of Arabidopsis (Klepikova *et al.*, 2015). *SOC1* and *AGL24* at the shoot apical meristem directly activate the expression of *LFY*, which alongside *AP1*, leads to flower development through the activation of genes including *AP3* (Lee and Lee, 2010).

The expression of these five genes over time in DH1012 is different to the expression profiles of their Arabidopsis orthologues. In Arabidopsis, *AP1, AP3, LFY* and *SOC1* expression increases over time, and whilst this is also true for DH1012, this increase is seen much later in development than it is in Arabidopsis. In contrast, *AGL24* expression in Arabidopsis increases early on in development before decreasing prior to the floral transition. In DH1012, the expression profile for *AGL24* is very different, after a reduction in expression prior to the floral transition, it seems to stabilise, something that is not seen in Arabidopsis.

Curve registration was used to compare the expression profiles of these homologous genes between Arabidopsis and DH1012 in more detail (Fig. 4.5). Following registration, varying results were observed. For *LFY*, the expression profiles between the species were superimposed onto one another, indicating similar, although desynchronised, expression patterns. As demonstrated through the table of optimal transformation function parameter estimates, the difference in gene expression profiles between the species for *LFY* can be explained by a combination of shift and stretch factors.

For *AGL24, AP1* and *SOC1*, the curve registration was able to superimpose the expression patterns between the Arabidopsis and DH1012 homologues, however variation was seen in regions of the alignment (Fig. 4.5). This signified the alignment of the expression profiles of these genes may not be as successful as it was for the *LFY* homologues. However, similarly to *LFY*, the differences in the gene's expression profiles can be explained by a combination of shift and stretch factors.

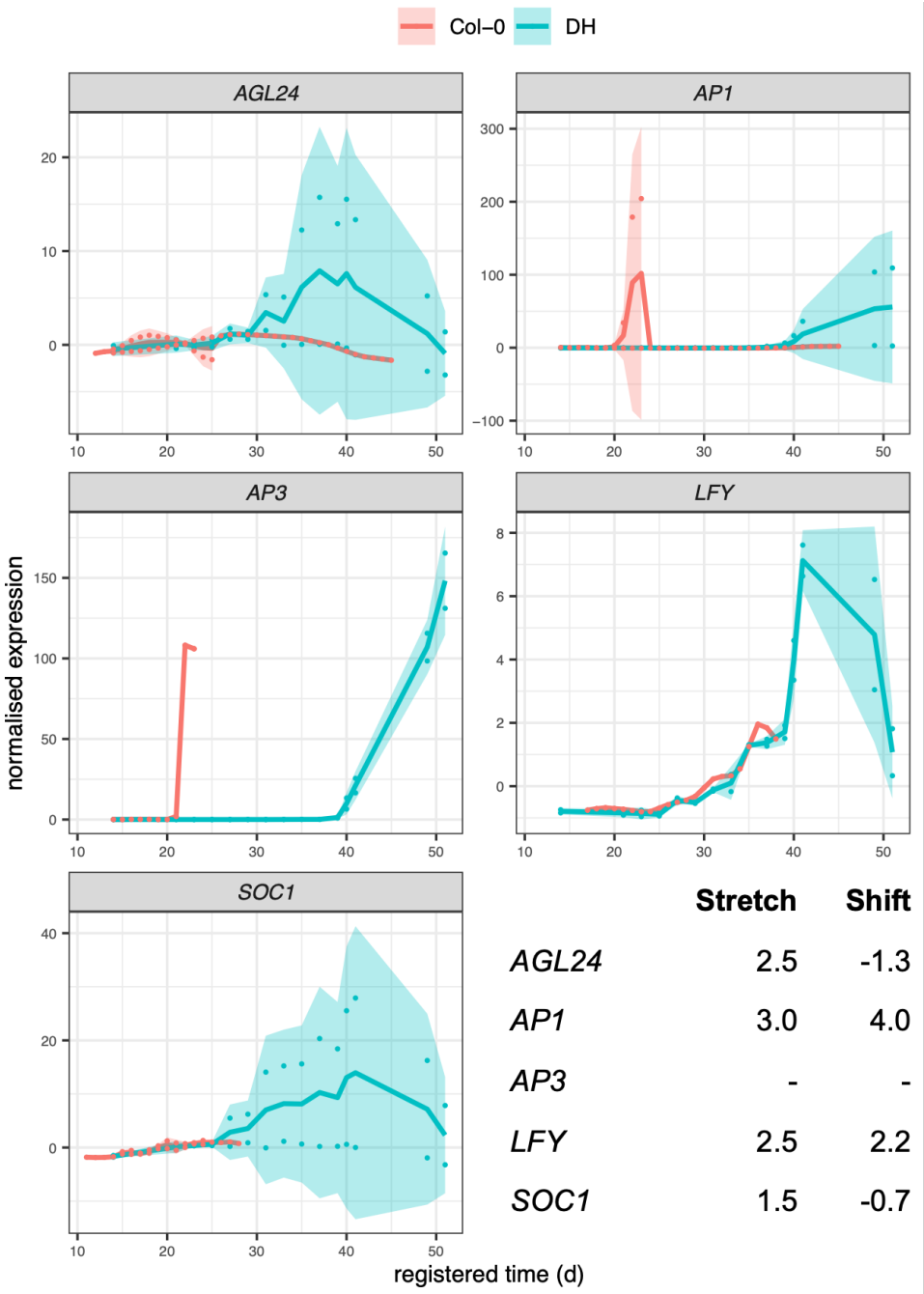


(taken from Calderwood *et al.* 2021)

Figure 4.4: The expression of five representative floral genes were investigated. A) *Arabidopsis thaliana* Col-0 and B) *Brassica oleracea* DH1012. Expression of DH1012 paralogues are summed. Morphologically identified floral transition is indicated by a vertical black line, 14 d in Arabidopsis and 35 d in DH1012. The floral transition and timings of gene expression changes relative to other genes, differ between Arabidopsis and DH1012.

Finally, for *AP3* we see that the Arabidopsis and DH1012 expression patterns have not been successfully superimposed through the curve registration (Fig. 4.5). The two appear to have very similar expression profiles, a period of very low-level expression followed by a significant increase in expression. However, for the Arabidopsis expression profile, the final timepoint is not an increase but instead a plateau. Furthermore, for *AP3* expression is only seen in the last three timepoints of the series. This plateau at the final timepoint, combined with minimal timepoints to describe the sudden change in expression within the Arabidopsis time series, are likely to have resulted in this lack of registration between the two.

For the majority of the genes investigated, the expression profiles between the Arabidopsis and DH1012 orthologues could be superimposed, confirming the hypothesis that expression of homologous genes are similar between the two species. Thus, the differences observed between the expression profiles between Arabidopsis and DH1012 can be largely explained through differences in timing rather than differences in expression dynamics.



	Stretch	Shift
AGL24	2.5	-1.3
AP1	3.0	4.0
AP3	-	-
LFY	2.5	2.2
SOC1	1.5	-0.7

Figure 4.5: Curve registration of five representative floral transition genes revealed for many of the genes their expression profiles are similar, but the timings are different.

The table displays the optimal registration parameters identified for each gene, stretch is the stretch factor applied and shift is the delay applied in days to the Arabidopsis data. The Arabidopsis *AP3* gene expression profile was unsuccessfully registered to the DH1012 expression profile.

4.3.4 Exploring the registration of individual gene paralogues reveals different expression profiles

To investigate the variation in the registered expression profiles for *AGL24, AP1* and *SOC1*, the expression profiles for the paralogues of these genes were plotted individually (Fig. 4.6) and the optimal registration parameters considered individually (Table 4.1). For *AGL24*, which has two copies in *B. oleracea,* the C07 paralogue could be registered, but the paralogue on C01 could not. For BO1G039080, the expression begins very low and is stable, but is then followed by an increase in expression around the floral transition and then a decrease in expression in the final stages of the time series. In contrast, expression of BO7G109590 increases in the lead up to the floral transition, before decreasing to a level lower than its initial level after the transition. This pattern of expression for BO7G109590, is also demonstrated in the Arabidopsis *AGL24* expression profile, hence it is able to be registered.

With *AP1*, three paralogues are identified in *B. oleracea*, one on C02 and two on C06. However, only the two C06 paralogues were expressed within the apex tissue in this study, therefore the C02 paralogue was not included in the curve registration. Of the two C06 paralogues, one registered successfully, but the other did not, unlike in the expression profiles of the Arabidopsis *AP1* and BO6G095760, in which expression increases over time. Within the expression profile of BO6G108600, a decrease in expression is seen at the final timepoint, which could account for the lack of registration between this copy and the Arabidopsis *AP1*.

Three paralogues of *SOC1* have also been identified in *B. oleracea*, one on C03 and two on C04 (Schiessl *et al.*, 2017) but as of yet no functional studies have been carried out on these paralogues. All three paralogues were expressed in the apex tissue for this time series experiment, however one of the C04 paralogues, BO4G195720, was expressed at a very low level and subsequently not included in the curve registration process. For the remaining two paralogues, both copies were registered successfully, however their optimal registration parameters differed. For both paralogues the optimal stretch value was the same, a stretch of 1.5, but the shift value differed between the two. BO3G038880 had a shift value of 0.9, whereas the C04 paralogue had a shift value of -3 (Table 4.1). This suggests that whilst the two paralogues have a similar expression profile, they are desynchronised from each other as well as from the Arabidopsis homologue.

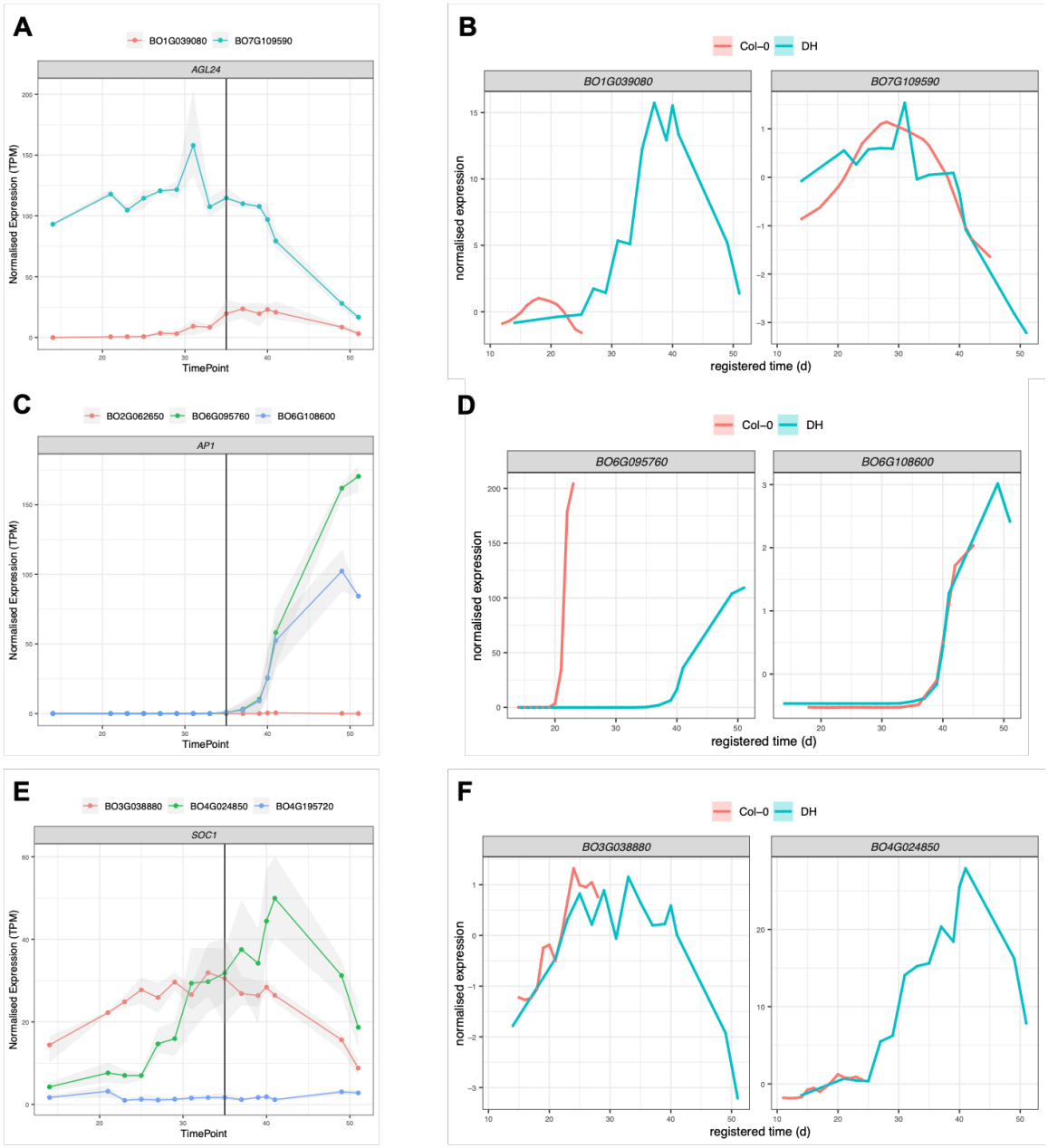


Figure 4.6: Investigation into gene paralogues of *AGL24, AP1* and *SOC1* revealed differences in the expression profiles between copies.

Expression profiles of individual paralogues of A) AGL24, B) AP1 and C) SOC1 over the time series. Curve registration of the paralogues individually, demonstrated differences between paralogues that would explain for the variation seen when registering the summed expression of the paralogues. B) BO1G039080 *AGL24* was not successfully registered to its Arabidopsis homologue, however BO7G109590 *AGL24* was. D) BO6G095760 *AP1* was not successfully registered to the Arabidopsis homologue, however BO6G108600 *AP1* was. F) Both paralogues of *SOC1,* BO3G03880 and BO4G024850, were successfully registered but had very different optimal registration parameters.

Table 4.1: Gene paralogue information and the optimal registration parameters used for each of the five representative floral genes.

CDS Model	Stretch	Shift	Symbol	Arabidopsis Identifier
BO1G039080	1.5	-2.22	AGL24	AT4G24540
BO7G109590	3.5	-0.44	AGL24	AT4G24540
BO6G095760	3	4.00	AP1	AT1G69120
BO6G108600	3	4.00	AP1	AT1G69120
BO4G120010	-	-	AP3	AT3G54340
BO8G083600	-	-	AP3	AT3G54340
BO2G161690	2.5	2.22	LFY	AT5G61850
BO3G109270	2.5	2.22	LFY	AT5G61850
BO3G038880	1.5	0.89	SOC1	AT2G45660
BO4G024850	1.5	-3.00	SOC1	AT2G45660

Plotting the paralogues individually demonstrated that the cause of the variation seen in some areas of the registered expression profiles, was due to differences between the expression dynamics of individual paralogues. These differences resulted in different optimal registration parameters between the paralogues and in one case a situation where one paralogue was registered and one was not, details which were masked by summing the expression of the paralogues.

4.3.5 Gene regulatory networks to further explore the differences between *AP1* paralogues

The successful registration of BO6G108600 *AP1* and not BO6G095760 *AP1* to the Arabidopsis *AP1* expression profile, opened a unique opportunity to explore the roles of these paralogues using gene regulatory network inference. The successful registration of BO6G095760 *AP1* suggests that it could be performing a similar role to the Arabidopsis *AP1,* as it is demonstrating a similar expression profile. Conversely, the inability to register BO6G108600 *AP1* and the subsequent difference in its expression profile suggests it may be performing a

different role to the Arabidopsis homologue. Causal structural inference (CSI) was used to generate networks for each *AP1* copy individually to test this.

Networks were generated for each *AP1* paralogue, focusing on two genes the Arabidopsis *AP1* homologue is known to interact with, namely *LFY* and *TFL1* (Serrano-Mislata *et al.*, 2017). *AP1, LFY* and *TFL1* interact with one another in a series of feedback loops and are key regulators of the floral transition (Fig. 4.6A). *AP1* and *LFY* are thought to act synergistically (Weigel *et al.*, 1992), promoting the expression of one another. *AP1 LFY* double null mutants exhibit a severe flowering phenotype. Furthermore, *AP1* and *LFY* act antagonistically to regulate *TFL1* expression, with *LFY* promoting *TFL1* expression, whilst *AP1* suppresses it (Goslin *et al.*, 2017; Serrano-Mislata *et al.*, 2017). It was hypothesised a similar network to that described within Arabidopsis would be generated for BO6G108600 *AP1*, *TFL1* and *LFY* due to similarities in expression dynamics. However, interestingly for both copies of *AP1*, the CSI algorithm generated the same network, which did not include *AP1* (Fig. 4.7B). This was unexpected, especially for BO6G108600 *AP1*, one explanation for the result could be that there are not enough timepoints within the *AP1* expression profile for the algorithm to infer a causal link, as expression is only seen at the later timepoints in the time series.

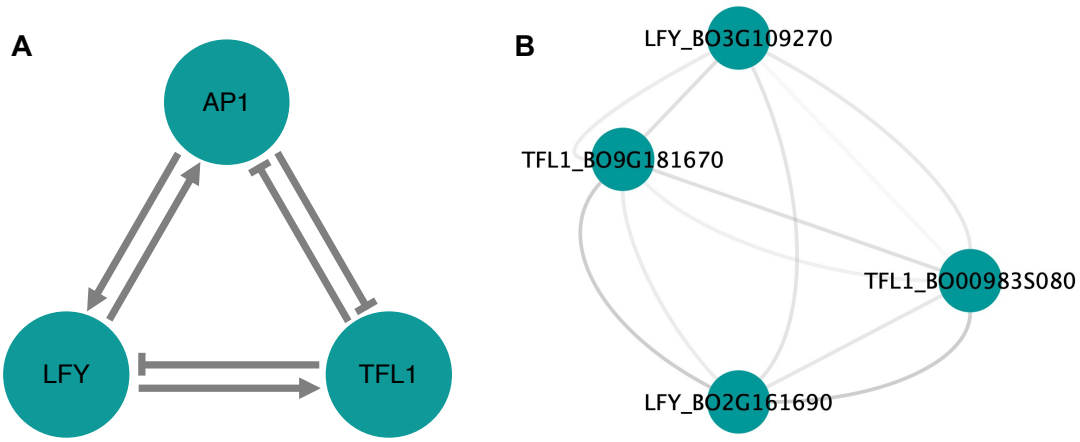


Figure 4.7: *AP1, LFY* and *TFL1* are tightly regulated in the model species *A. thaliana*. A) Schematic representation of the regulatory interactions between *AP1, LFY* and *TFL1* in Arabidopsis, according to current literature. B) CSI gene regulatory networks between expressed paralogues of *AP1, LFY* and *TFL1* in *B. oleracea* DH1012. *AP1* is not included in the network, suggesting there is insufficient data points for the CSI algorithm. Transparency of edges is relative to the edge weights. Network visualised using Geneious.

4.3.6 *TFL1* expression dynamics are similar between paralogues in DH1012

TFL1 is a floral repressor gene and within *B. napus*, expression profiles of paralogous *TFL1* gene copies have demonstrated regulatory differences (Jones, 2017). Within *B. oleracea*, a close relative of *B. napus*, the magnitude of expression differs between paralogues, but the overall expression profile is very similar. Both paralogues identified were able to be successfully aligned to the Arabidopsis *TFL1* expression profile using gene registration (Fig. 4.8). The optimal registration parameters for both were the same, a stretch of 2.5 and no shift (Table 4.2). This confirms that the expression profile of each paralogue differs in magnitude, but not in pattern and that this pattern is similar to that of the Arabidopsis *TFL1*. This suggests that, unlike in *B. napus,* the *B. oleracea TFL1* paralogues are performing a similar role both to one another and to their Arabidopsis homologue.

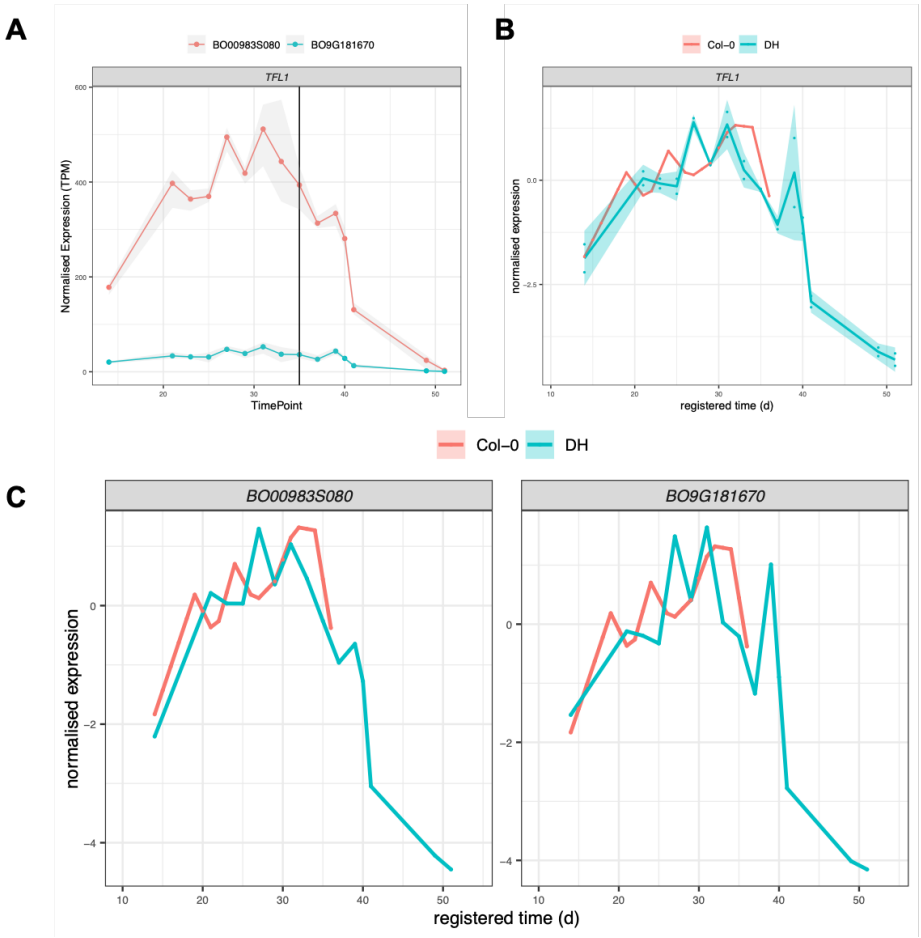


Figure 4.8: Previous work in *B. napus* revealed differences in the paralogues of *TFL1*, therefore the *B. oleracea TFL1* paralogues were investigated.

A) Normalised expression using TPM, of *TFL1* paralogues identified in DH1012 using BLAST analysis. B) Curve registration was employed and revealed that the Arabidopsis *TFL1* expression profile could be successfully superimposed over the DH1012 *TFL1* expression profile. C) Curve registration of the individual *TFL1* gene paralogues revealed both could successfully be aligned the Arabidopsis *TFL1* expression profile and that their optimal registration parameters were the same. This indicates that the difference between the expression profiles of the two paralogues is in magnitude and not due to a difference in expression pattern.

Table 4.2: Gene paralogue information for *TFL1* and the optimal registration parameters calculated for each.

CDS Model	Stretch	Shift	Symbol	Arabidopsis Identifier
BO00983S080	2.5	0	TFL1	AT5G03840
BO9G181670	2.5	0	TFL1	AT5G03840

4.3.7 Using gene registration to explore a role for *FLC* in DH1012

FLC is a well-studied floral repressor (Schranz *et al.*, 2002; Okazaki *et al.*, 2007; Razi *et al.*, 2008; Irwin *et al.*, 2016) and five orthologues of Arabidopsis *FLC* have been identified in *B. oleracea* (Okazaki *et al.*, 2007). *FLC* is known to play a key role in the vernalisation pathway and is thought to control flowering in a dosage dependent manner, as the number of *FLC* gene copies present appears to have an effect on flowering time. DH1012 is a rapid cycling *B. oleracea* variety, requiring no vernalisation to flower, however expression of all four *FLC* copies were expressed in the apex tissue across the time series (Fig. 4.9A). For two of the four copies, BO9G173400 and BO9G173370, their expression was downregulated over the time series, and this downregulation occurred prior to the meristem switching from a vegetative to inflorescence state at 35 d, despite the lack of cold exposure to induce this. Furthermore, the expression profiles of these three copies were successfully registered to their Arabidopsis homologue (Fig. 4.8C) suggesting they are performing a similar function.

The remaining *FLC* paralogue, displays a different expression pattern. BO3G005470 is upregulated prior to 35 d, after which its expression decreases and then seems to stabilise. This is the opposite of what we would expect from a floral repressor gene, suggesting in DH1012 it may be carrying out a different function, or because of the dosage dependent manner in which the *FLC* copies appear to function, perhaps its expression is outweighed by the remaining paralogues.

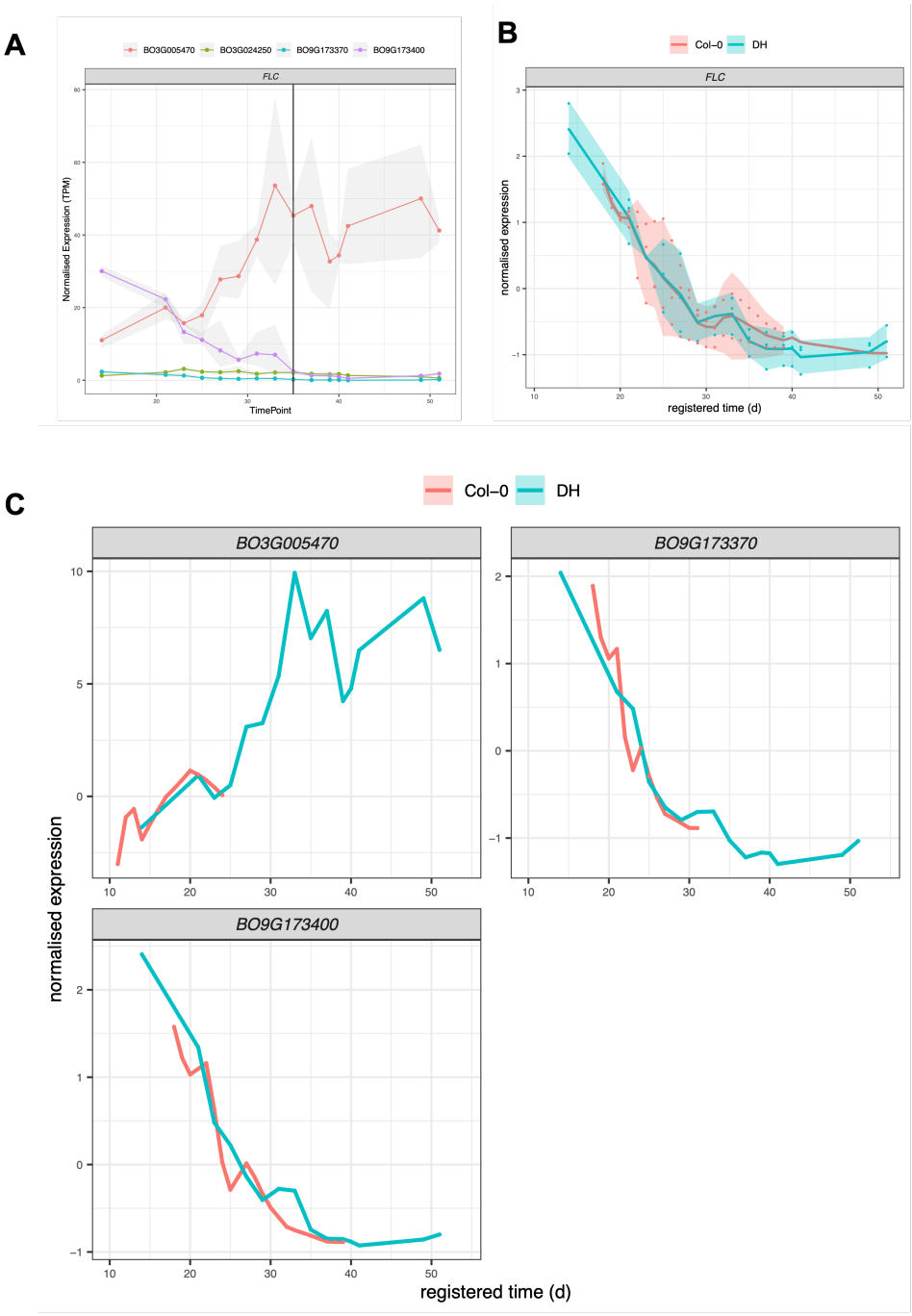


Figure 4.9: *FLC* is expressed in the non-vernalisation requiring line DH1012.

A) Normalised expression using TPM, of *FLC* paralogues in DH1012. B) Curve registration was employed and revealed that the Arabidopsis *FLC* expression profile could be successfully superimposed over the DH1012 *FLC* expression profile. C) Curve registration of the individual *FLC* gene paralogues revealed all could be registered, but that different optimal registration parameters were required for each.

Table 3.3: Gene paralogue information for *FLC* and the optimal registration parameters used for each.

CDS Model	Stretch	Shift	Symbol	Arabidopsis Identifier
BO3G005470	1.5	-3.0	FLC	AT5G10140
BO9G173400	2.5	3.1	FLC	AT5G10140
BO9G173370	1.5	4.0	FLC	AT5G10140

4.4 Discussion

B. oleracea and Arabidopsis are related species, both being members of the *Brassicaceae* family, and they exhibit similar apex morphologies over development. Despite these similarities, the expression profiles of many key flowering time genes differ between the two species. However, through curve registration we demonstrate that this difference can partly be explained by the desynchronisation of gene expression between the two species, a finding that is consistent to the differences seen between Arabidopsis and species *B. rapa* (Calderwood, Hepworth, *et al.*, 2021). Furthermore, through GO enrichment it was observed that genes with similar registration parameters are involved in similar biological processes. Functional processes likely act in concert, but this result indicates that different processes have likely become desynchronised between Arabidopsis and DH1012. This is in contrast to an overall change that might be expected between the two, due to the similarities seen in their morphological development.

Although gene registration demonstrated that expression differences between Arabidopsis and DH1012 can largely be explained by differences in timing and magnitude of expression, rather than differences in expression patterns, several paralogues were found to exhibit different expression dynamics from one another, and it was necessary to consider them individually during the registration process.

AGL24 is an Arabidopsis floral meristem identity gene (Gregis *et al.*, 2008, 2009; Liu *et al.*, 2008) that encodes a MADS-box transcription factor. *AGL24* works closely with another MADS-box gene, *SOC1,* in a positive feedback loop to directly regulate *LFY* expression. In Arabidopsis, *AGL24* expression is initially upregulated by vernalisation independently of *FLC* (Michaels *et al.*, 2003), however it has been shown that *AGL24* expression must then be repressed, as continued *AGL24* expression can cause floral reversion. This pattern of expression was seen in the Arabidopsis data and for the C07 paralogue of *AGL24* in DH1012

(Fig. 4.6A). However, expression of the C01 paralogue increased after the meristem had switched to producing floral primordia, hence the inability to superimpose this expression profile over the Arabidopsis expression profile using curve registration. This pattern of expression is in complete contrast to what was expected from knowledge in Arabidopsis and suggests that this *AGL24* paralogue may be performing a different function within DH1012, although this can only be speculative without follow up with functional analysis. Furthermore, it is interesting that this gene is upregulated by vernalisation in Arabidopsis and yet despite the lack of vernalisation requirement and exposure in DH1012, upregulation of *AGL24* is still observed. Again, this difference to its Arabidopsis homologue suggests, that *AGL24* could have developed a different function within DH1012. With this delayed increase in expression relative to its Arabidopsis homologue, perhaps this paralogue has maintained a role in flowering but is involved in a later process, such as floral organ development.

SOC1 in Arabidopsis performs a critical role in the flowering pathway, acting as an integrator for signals from the photoperiod, temperature, hormones and ageing pathways (Lee and Lee, 2010). Whilst three paralogues of *SOC1* have been identified in *B. oleracea*, no functional studies have yet been conducted (Leijten *et al.*, 2018). Here we show that in DH1012 only two of the three paralogues are expressed in the apex over the floral transition. Of the two that are expressed, both are present on C04 but have different expression dynamics, which is reflected in the different optimal registration parameters required to align both to the Arabidopsis *SOC1* expression profile, as was seen with *AGL24*. The C03 paralogue of *SOC1* peaks in its expression at 43 d, which is after the meristem had become an inflorescence, however the peak in expression for its Arabidopsis homologue occurs prior to the meristem becoming an inflorescence (Fig. 4.4A). This could suggest that this paralogue has greater importance later in the floral transition relative to its Arabidopsis homologue.

AP1 is a floral identity gene and determines sepal and petal development through encoding a putative transcription factor (Gustafson-Brown, Savidge and Yanofsky, 1994). Unlike *AGL24*, functional analysis of the *B. oleracea AP1* paralogues has been carried out (Lowman and Purugganan, 1999; Smith and King, 2000; Kop *et al.*, 2003). Three paralogues of *AP1* have been identified, *BoAP1-a* (Anthony, James and Jordan, 1996), *BoAP1-b* (Lowman and Purugganan, 1999) and *BoAP1-c* (Smith and King, 2000). It has been observed that *BoAP1-b* is non-functional, which accounts for the lack of expression seen here in the C02 paralogue of *AP1* in DH1012, resulting in its exclusion from the curve registration. Both *BoAP1-a* and *BoAP1-c* have been located on the C06 chromosome using QTL mapping (Hasan *et al.*, 2016), this region of the C06 chromosome within *B. oleracea* shows collinearity to Arabidopsis C01, which is known to contain key flowering time genes. For the remaining two *AP1* paralogues,

their expression patterns are largely similar to the Arabidopsis *AP1* expression pattern, a period of no expression followed by a dramatic increase in expression. However, for one of the C06 paralogues this is followed by a decrease, which is not seen in the Arabidopsis expression profile, and has led to the two not being successfully registered. It is possible that the observed decrease in BO6G10860 *AP1* is compensated for by the increase in BO6G095760 *AP1* and that this change in dynamics is consistent with the gene dosage hypothesis (Papp, Pál and Hurst, 2003; Conant and Wolfe, 2008).

For *AGL24, SOC1* and *AP1*, such differences between the gene expression patterns of the paralogues raise many questions as to their functions within DH1012 in comparison to what is known of their homologues within Arabidopsis. Follow up to this investigative work, in the form of functional analysis of these paralogues, could aid identification of function. Generating Arabidopsis knockouts for each of these genes and complementation with the paralogues, would determine which paralogues are performing the same role as their Arabidopsis homologue, however it would not provide answers as to what any novel functions that have arisen may be. Whilst there wasn't the scope to perform such functional analysis within the project, gene regulatory network inference was used to try and understand more about how these genes were interacting.

Network inference was used to build gene regulatory networks for each of the expressed *AP1* paralogues in DH1012. The interaction between *AP1, LFY* and *TFL1* in Arabidopsis is well documented and known to play an important role in meristem identity (Bowman *et al.*, 1993; Liljegren *et al.*, 1999; Goslin *et al.*, 2017; Serrano-Mislata *et al.*, 2017). If the interactions between BO6G095760 *AP1* and the *LFY* and *TFL1* paralogues mirrored what is seen in Arabidopsis, this would be further evidence supporting that this *AP1* copy has a conserved role within DH1012. However, neither copy of *AP1* was included in the respective gene regulatory networks. This may be attributed to the fact that both copies are not expressed until 37 d, meaning there are only five data points for the CSI algorithm to work with in both cases. This is a limitation of both the method and the data available, perhaps with an increased number of timepoints it would be possible to generate gene regulatory networks for the *AP1* paralogues. It should also be noted that validation has yet to be carried out for these networks so they must be considered exploratory. In the future it would be useful to try and replicate the Arabidopsis *AP1, LFY, TFL1* regulatory network with the same method to check the efficacy of the method.

Investigation was carried out into the floral repressor *FLC. FLC* has been studied in *B. oleracea,* and its role in the vernalisation pathway as part of the floral transition confirmed

(Schranz *et al.*, 2002; Pires *et al.*, 2004; Lin *et al.*, 2005; Okazaki *et al.*, 2007; Razi *et al.*, 2008; Irwin *et al.*, 2016). Two of the four *FLC* copies identified in DH1012 are downregulated at the apex prior to 35 d, at which time the apex was seen to have transitioned from a vegetative to an inflorescence meristem. As a floral repressor, *FLC* expression needs to be suppressed for flowering to occur, however it is unclear how this is being mediated in DH1012 without the cold exposure to induce this. Furthermore, the remaining copy, BO3G005470, is upregulated prior to 35 d, which is in complete contrast to the expression pattern we would expect to see if the copy was functioning as a floral repressor. This suggests it could be performing a different role within DH1012. This hypothesis is likely, because we know that the C02 copy of *FLC*, which has been identified in other *B. oleracea* (Razi *et al.*, 2008; Ridge *et al.*, 2015; Irwin *et al.*, 2016), is not present in DH1012 and other rapid cycling lines, such as TO1000, and this is largely thought to account for their rapid cycling behaviour. It is interesting then that we are seeing expression of an *FLC* copy in the apex prior to vernalisation, as we would expect *FLC* to have little influence over flowering in DH1012. Functional investigation into the role of this *FLC* paralogue could help to provide an answer for the function behind this copy and the reason for its expression in the lead up to the switch from a vegetative to inflorescence meristem.

Chapter 5

Discussion

5.1 Chapter Summaries

5.1.1 Phenotyping and associative transcriptomics for Brassica vegetables

Here in the UK, we are experiencing warmer weather and more variable winter temperatures. The winter of 2020 was the fifth warmest for the UK in a series from 1884, and was also one of the wettest with an absence of frosts and snowfall due to the temperatures (Kendon *et al.*, 2021). Many *B. oleracea* crop vegetables require a period of cold exposure, known as vernalisation, for their correct development. The warmer temperatures seen in the UK mean that many of the crop vegetables produced here, especially strong vernalisation requiring types such as Brussels Sprouts, are at risk. To future proof the production of these vegetables we need to identify or produce lines which require little or no vernalisation.

Here we provide extensive phenotyping of a previously un-phenotyped subset of the *B. oleracea* Diversity Fixed Foundation Set, from the University of Warwick (University of Warwick, 2009). These phenotyping results are a valuable resource for the Brassica community. In determining the flowering times of these lines and the optimal conditions for that floral transition to take place, informed work can now be carried out on selected lines to determine the genes underlying the different flowering phenotypes observed. This phenotyping data includes detailed insights into the vernalisation requirements of each line and has enabled the identification of lines that may perform better under warmer winter conditions, and lines which require no vernalisation at all to produce the desired phenotype. This information could prove useful to the plant breeding community in the search for lines that are better adapted to the changing climatic conditions, including the warmer and more variable winter temperatures we are seeing here in the UK.

In addition to the identification of lines that are better adapted to changing climatic conditions, plant breeding efforts must also be focused on the production of new lines that are better suited

to this altered climate. Determining which genes underly phenotypic traits is a key step to achieving this goal. Associative transcriptomics is a powerful approach for the identification of molecular markers linked to trait controlling loci (Rafalski, 2010) and has been used for this purpose in numerous crop species (Harper *et al.*, 2012; Huang *et al.*, 2012; Romero Navarro *et al.*, 2017; Raman *et al.*, 2019). Here we present a novel associative transcriptomics pipeline for *B. oleracea*, which has been validated using crop type and phenotypic information on heading and flowering. Using this pipeline, we identify candidates for the vernalisation response in *B. oleracea* that are orthologues of known Arabidopsis floral regulators. The ability to identify such candidates is an important step and could provide a foundation for future breeding strategies. The development of this pipeline and its use in identifying candidates of complex traits, such as flowering time, demonstrate that this pipeline can be used for other traits of agronomic importance, such as disease resistance and germination, making it an invaluable resource.

5.1.2 Investigating the genetic basis of the floral transition in DH1012 using transcriptomics

B. oleracea is a hugely important crop species and is widely grown here in the UK. It is a member of the *Brassicaceae* family, however unlike in other cultivated members of the family, such as *B. napus*, much remains unknown about the floral transition in this species. Work has previously been carried out to investigate the time to curd specifically in cauliflower (Hand and Atherton, 1987; Anthony, James and Jordan, 1996; Matschegewski *et al.*, 2015; Hasan *et al.*, 2016; Rosen *et al.*, 2018), but few studies have been published on flowering time in other *B. oleracea* subspecies. Here we present the first study in *B. oleracea* to follow the transcriptome through the floral transition. We explore the roles of both the leaf and apex tissues in detail, examining the expression profiles of key genes of interest and looking at how they cluster together. This data is a highly useful resource and the first of its kind within the species, allowing us to track global changes in gene expression over the transitions.

Differential gene expression analysis was used to determine differences between the leaf and apex tissues. Overall, more genes were differentially expressed in the leaf tissue than the apex tissue. This is thought to be due to the fact that leaves age independently and therefore toward the end of the time series signs of ageing and senescence would be seen in the leaf tissue. Indeed, this was the case and GO enrichment analysis revealed enrichment of the terms GO:0007568 ageing and GO:0008219 cell death in the final leaf sample.

To identify genes that regulate the floral transition in *B. oleracea*, we compared the transcriptional profiles of three developmental stages across the time series; 14 d when the plants were vegetative, 35 d after the meristem had begun to exhibit floral primordia and 51 d when buds were visible at the apex. Through differential gene expression analysis we demonstrate a critical point in flower initiation is seen in DH1012, which mirrors observations previously made in Arabidopsis (Klepikova *et al.*, 2015). A marked increase in the number of both up and downregulated DEGs is seen at the point where the meristem transitions from vegetative to inflorescence. GO enrichment analysis revealed that these increases in DEG numbers coincide with the expression of genes involved in meristem identity, however these genes do not include homologues of many of the key players in meristem identity that have been identified from work within Arabidopsis. This suggests that there is a different core set of meristem identity genes at work in DH1012, relative to in Arabidopsis Col-0.

To assess trends in the data, self-organising maps (SOMs) were used to cluster gene expression over time. A 3 x 3 SOM was generated for the DEGs identified in the apex tissue. Cluster 6 of this SOM was of particular interest, as it included many orthologues of Arabidopsis floral regulators. Within this cluster, three *FD* paralogues, and paralogues of *SOC1, AGL24, TOE3* and *SPL4* were found. This indicates that despite the differences in meristem identity genes, in DH1012 many homologues of Arabidopsis floral regulators exhibit similar gene expression profiles to their Arabidopsis counterparts. SOM analysis also identified three key expression patterns which occurred most frequently across both tissues. The first was a rapid decline in expression at the first few timepoints, followed by low expression for the remainder of the time series. If a gene was exhibiting this expression profile it would be expected that it would play a role in the early vegetative stage. The second profile was low expression followed a large increase in expression toward the final few timepoints, in which bud emergence occurred. This profile suggests the corresponding genes would be involved in floral organ identity. Finally, a marked increase in expression levels at the middle of the time series, the time at which the apex transitioned from a vegetative state to an inflorescence meristem, indicating the onset of the expression of genes involved in meristem identity. These findings mirrored the conclusions obtained from patterns in DEG number across the time series.

This investigation into the leaf and apex tissues in DH1012 lays the foundations for targeted work on specific floral regulators in DH1012. Key candidates have been identified, such as *FD* paralogues, which with further functional analysis could provide key insight into the control logic of flowering in DH1012. This targeted work will enable us to gain a greater

understanding of the floral transition in *B. oleracea* and begin to describe where the differences lie between DH1012 and Arabidopsis.

5.1.3 Using comparative transcriptomics to gain a better understanding of the floral transition in *B. oleracea*

The study of flowering has been dominated by work on the model species *Arabidopsis thaliana*. This has generated a demand for methodology to allow the translation of this information into other species, with a specific focus on the translation into crop species. Using a method previously employed in the close relative of *B. oleracea, B. rapa*, we have been able to explain seemingly large differences in the expression profiles of DH1012 and Arabidopsis over the floral transition. Through the application of this gene registration technique, we have been able to align the expression profiles of many key floral regulators between the species. This indicates that differences in gene expression profiles between the two can largely be explained by a difference in timing or magnitude, as opposed to a difference in expression dynamics.

However despite the ability to align the majority of genes to their Arabidopsis homologues, there was no universal set of optimal registration parameters that could be used to describe the differences. This suggested that there is not a linear progression in gene expression states between the two species. However, groups of genes with the similar registration parameters were found to be involved in similar biological processes. This finding indicates that different developmental processes occur at different rates between Arabidopsis and DH1012, which is interesting, as morphologically the difference in the development of the shoot apical meristem appears to be linear (Fig. 4.2). This means that any analysis that tries to compare between these related species may risk oversimplification, for example just comparing raw gene expression profiles. Whilst the findings of this work demonstrate there is no experimentally easy way to compare Arabidopsis and DH1012, they do show that comparisons can be made.

Here we have made use of this methodology to explore orthologues of known Arabidopsis floral regulators, however this method could be extended to explore other important biological processes such as disease response or germination between species. This is a valuable tool to have, however it is worth noting that it is important to consider gene paralogues individually, rather than just as a summed entity. We have demonstrated here that there are often marked differences in gene expression profiles between homologues, which is masked if their

expression is summed. Taking individual paralogue dynamics into account will allow us to gain a greater understanding of the genetic regulation of many biological processes.

5.2 Outlooks and Limitations

Obtaining sufficient material to generate a panel large enough for associative transcriptomics in *B. oleracea* was the first challenge. Due to the self-incompatibility that is widespread throughout *B. oleracea*, generating enough seed to carry out a large-scale experiment such as this is a difficult task. Consequently, we knew we would be working with fewer lines than is usually used in associative transcriptomic studies. To tackle this, we generated a population structure specific to the subset of lines used in the phenotyping experiment and carried out extensive validation of this population structure. In an ideal world, it would be beneficial to include more lines to this panel, with a particular focus on the inclusion of more landraces to help mine for more trait-controlling loci.

Juvenility is an important, but largely uncharted topic in Brassicas. Here we investigated the effects of giving a six- or ten-week pre-growth period prior to vernalisation exposure. We were able to identify lines which appear to have a juvenile phase of six weeks or less, as they were able to respond to vernalisation and flower, unlike many others which were unable to respond to vernalisation after the shorter pre-growth. However, due to the intrinsic link between juvenility and vernalisation, it was difficult to determine the length of the juvenile phase for any of these lines. For a plant to flower it must be both in the adult phase and have its vernalisation requirement saturated. Consequently, the optimal vernalisation length and temperature for the line in question needs to be determined and saturated, before investigation into juvenility, to be certain that any effect measured is a consequence of the pre-growth period alone. An experiment was planned using some of the candidate lines for juvenility identified here. It was planned that candidate lines would be exposed to pre-growth periods of 1 – 6 weeks, at weekly increments, before vernalisation at 5 °C for 12 weeks, to ensure saturation of any vernalisation requirement. The aim was to narrow down the window for the juvenile-to-adult transition in these lines. However, time constraints for the project meant this work couldn't be completed. Looking to the future, if this experiment was to be carried out, I think it would provide us with more specific knowledge on juvenility in a subset of *B. oleracea* lines, which could then become useful tools in searching for the genes which control this transition.

Both GWAS and GEM analysis were used to identify candidates involved in flowering time in *B. oleracea*. The 110555 SNPs used for the GWAS were identified using transcriptomics. This means that only SNPs from coding regions will be included in the analysis and consequently, a host of SNPs will not have been included compared to if we had used traditional genomic SNP identification (Huang *et al.*, 2012). Despite the larger number of SNPs that can be identified using whole genome resequencing, the use of transcriptome resequencing using mRNA-seq, means transcript abundance data was available to us and GEM analysis could subsequently be conducted. Nevertheless, repeating the construction of this pipeline for GWAS using genomic SNPs could lead to the identification of more associations and provide a more complete picture.

Further to this, the transcriptome resequencing, used to generate the SNPs for GWAS and expression data for GEM analysis, was carried out on 14 d leaf tissue. The tissue chosen and the timepoint selected, may have again limited the available SNPs within the panel, as many floral genes may not be being expressed at this time and/or in this tissue. Using our transcriptomic data generated for the rapid cycling line DH1012, we have shown that at 14 d the plant is vegetative, therefore this timepoint may be too early to capture SNPs in flowering time associated genes within many *B. oleracea* lines, as it may be too early to detect the expression of many genes associated with the floral transition. If transcriptome resequencing was to be used again to generate SNPs, perhaps it would be more appropriate to use apex tissue sampled at a later timepoint, to potentially identify a greater number of associations with the flowering traits analysed here.

Despite these drawbacks, candidates were successfully identified using the associative transcriptomics pipeline, including *miR172D*. This candidate was identified for two distinct traits using GWAS analysis, days to BBCH60 under non-vernalising conditions and the difference in days to BBCH51 between six and twelve weeks of vernalisation at 15 °C. The homologous miR172 family in Arabidopsis is well documented for its role in the ageing pathway as part of the floral transition (Jung *et al.*, no date; Wu *et al.*, 2009; Zhu and Helliwell, 2011; Spanudakis and Jackson, 2014; Teotia and Tang, 2015). In Arabidopsis, the miR172 family is responsible for post-transcriptional suppression of *APETALA-1*-like genes, including the *TARGET OF EAT* family, which promotes the floral transition (Aukerman and Sakai, 2003). Functional analysis of *miR172D* in *B. oleracea* is the next step in confirming its role in the floral transition in this species. A transformation platform is available at the John Innes Centre for DH1012. Therefore, it would be possible to generate DH1012 plants that are knockouts for *miR172D*. In generating knockouts, the phenotypic response could then be

monitored through scoring for days to BBCH51 and days to BBCH60, in comparison to wild type DH1012. Another option would be to perform a complementation experiment within Arabidopsis in which miR172 has been knocked out. This would determine whether the role of *miR172D* in *B. oleracea* is the same as the role of miR172 in Arabidopsis and is able to complement the mutant.

As well as associative transcriptomics, RNA-seq analysis was also a large part of this work. The initial aim was to compare DH1012 to a late flowering *B. oleracea* line, that had a strong vernalisation requirement. In choosing two lines with opposing life history strategies, it was hoped it would be possible to compare the genes involved in the floral transition between the two, to create a more representative view of the floral transition in *B. oleracea* as a whole. Two lines were initially selected for growth as the potential late flowering line, a cabbage GT090058 and a cauliflower GT090341. Non-uniform heading was observed through the cauliflower population, indicating that this line was not homozygous and therefore, it was not used for transcriptomic analysis. The cabbage line experienced severe nutrient deficiency and it is thought subsequently entered a state of dormancy and so was also not used for transcriptomic analysis. Over a year later the plants had still not gone to head, therefore a second growth using an alternate line was planned.

Using the information provided by the phenotyping experiment, I chose a broccoli line to grow. This broccoli, GT120210, was chosen for two reasons, firstly because broccoli has been selected for bud tissue, therefore growing it to BBCH51 made more sense than a leafy type such as cabbage and secondly because it was a late flowering line with a strong vernalisation requirement, requiring a 10-week pre-growth period followed by 12 weeks of vernalisation at 5 °C to go to head, a complete contrast to DH1012. Unfortunately, once again the plants experienced a severe nutrient deficiency following vernalisation. Looking forward, the potential for compare DH1012 to a later flowering *B. oleracea* line could provide very useful insights into the floral transition in *B. oleracea*, however a line should be chosen where much is known about its flowering phenotype and the nutrient levels should be thoroughly monitored throughout the course of the experiment.

Further to this, for the RNA-seq experiment, samples were taken of the whole DH1012 apex. For Arabidopsis, we know that there are distinct regions of the apex that consist of small subsets of cells with transcriptionally distinct regions (Laufs *et al.*, 1998; Yadav *et al.*, 2009). Sampling the apical meristem as a whole was suitable to enrich for apically expressed genes but may have resulted in a loss of transcriptional differences between the distinct regions within the apex. Laser microdissection has previously been used in maize (Brooks *et al.*, 2009)

and rice (Harrop *et al.*, 2016) apices to separate meristem domains and therefore perhaps could also be used in *B. oleracea* to resolve these transcriptional domains and give a more detailed understanding of gene expression at the apex.

Following transcriptomic analysis, gene registration was used to compare gene expression between Arabidopsis and DH1012. The Arabidopsis time series data used for comparison (Klepikova *et al.*, 2015) only sampled for apical tissue and as a consequence, only the apex tissue could be compared between Arabidopsis and DH1012. Whilst we know that the apex is a key site in the floral transition, it is known from work in the model species Arabidopsis that other tissues also play an important role. For example, the flowering time regulator *FT* is known to be expressed in the vasculature of cotyledons and leaf tissue (Abe *et al.*, 2005). Consequently, the expression of *FT* cannot be fully investigated using this method due to the lack of leaf tissue data in Arabidopsis. Looking forward, conducting gene registration on the leaf tissue would ensure the expression dynamics of more key floral regulators are able to be investigated. To do this, a comparative transcriptome time series experiment for Arabidopsis would need to be set up to sample leaf tissue.

Gene registration offers a robust method for the comparison of gene expression dynamics between species. Whilst so far it has only been used to compare to the model species Arabidopsis, there is no reason why it could not be used to compare between Brassicas themselves. For example, with the *B. rapa* time series data presented in Calderwood *et al.* 2021 (Calderwood, Hepworth, *et al.*, 2021), it would be possible to compare gene expression patterns between *B. rapa* R-O-18 and the *B. oleracea* DH1012 data presented here. This comparison could prove very useful in understanding how the two came together to produce *B. napus*. Furthermore, as both lines are rapid cycling varieties of their respective species, requiring no vernalisation to flower, any differences between the two could be attributed to differences between species as opposed to differences in their life history strategies. This comparison could also help to decipher which paralogues are behaving in similar ways between the two and which may have diverged. Moreover, using time series data generated by the BBSRC funded BRAVO consortium, it would be possible not only to compare between species, but also to explore the differences within a species which could help to decipher key genetic components that confer different life history strategies. For example, transcriptome time series data is available for *B. napus* winter oilseed rape line Express and also for the spring oilseed rape variety Stellar, using gene registration between the two could help to determine what the key genes are that cause the difference in their life history strategies.

Through gene registration it was also possible to make clear differences between the gene expression profiles of gene paralogues. Specifically in the successful registration of BO6G108600 *AP1* and not BO6G095760 *AP1* to the Arabidopsis *AP1* expression profile. Exploratory gene regulatory networks were developed for the *AP1* paralogues individually, to look at the regulatory links between the paralogues and paralogues of *LFY* and *TFL1*, homologues of which are known to interact with the Arabidopsis *AP1*. It was hypothesised for BO6G108600 *AP1*, the network generated would resemble that for Arabidopsis *AP1*, however this was not the case. The networks for both *AP1* paralogues were identical and neither contained *AP1*. The networks were generated using expression data, but expression for both *AP1* copies was only seen after 37 d meaning there were only five time points for the CSI algorithm to work with. This is a low number and could be the reason behind the exclusion of *AP1* from the networks. Additional timepoints in the time series after 51 d would be needed to verify this conclusion.

The CSI algorithm itself is a rapid way of generating gene regulatory networks to describe gene interactions. It was designed to reconstruct regulatory models from time series data and therefore was a fitting option for the type of data used here. However, CSI requires prior knowledge of what interactions may be occurring, as it needs the user to input gene expression data for specific genes. In this regard it is good for verifying existing hypotheses, but perhaps the application of another method for network inference could help to uncover novel gene interactions to describe flowering time in *B. oleracea*.

5.3 Concluding Remarks

The aim of this project was to gain a better understanding of the genetic control of flowering time in *B. oleracea,* and this has been achieved through the use of associative transcriptomics and RNA-seq analysis.

Here we provide a novel associative transcriptomics pipeline for *B. oleracea,* which has been validated and used to determine several candidates for the floral transition. This work also provides the first transcriptomic analysis of the floral transition in *B. oleracea*. Analysis of leaf and apex tissue has identified a peak in DEG numbers that coincides with the switch from a vegetative to inflorescence meristem. Furthermore, gene registration between Arabidopsis and DH1012 revealed that largely gene expression dynamics are similar between the two species, but differently synchronised. This suggests that, as hypothesised, the gene regulatory network for flowering is likely to be similar in DH1012 to Arabidopsis. However, key

differences were seen between some gene paralogues in DH1012, which indicates that some novel functions may have been developed for different gene copies.

This thesis provides insights into some of the key regulators of flowering time in *B. oleracea*, and work is underway to determine a gene regulatory network to describe the floral transition in *B. oleracea*. Given the genome complexity, the elucidation of such a network has proven to be challenging and this project highlights some of the obstacles for a crop that contains multiple gene paralogues that need to be overcome. We have made advances, but more work is needed to understand the differences between the gene paralogues present in *B. oleracea*, not only at an expression level, but also in their protein activity. Building on the current findings provides a clear direction for future work on flowering time in *B. oleracea* in light of changing climatic conditions.

References

Abe, M. *et al.* (2005) 'FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex', *Science*. Springer-Verlag, 309(5737), pp. 1052–1056. doi: 10.1126/science.1115983.

Alejandra Mandel, M. *et al.* (1992) 'Molecular characterization of the Arabidopsis floral homeotic gene APETALA1', *Nature*. Nature Publishing Group, 360(6401), pp. 273–277. doi: 10.1038/360273a0.

Alvarez, J. *et al.* (1992) 'Terminal flower: a gene affecting inflorescence development in Arabidopsis thaliana', *The Plant Journal*. John Wiley & Sons, Ltd, 2(1), pp. 103–116. doi: 10.1111/j.1365-313X.1992.00103.x.

Anders, S. (2010) *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*, *Soil*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 8 July 2021).

Andrés, F. and Coupland, G. (2012) 'The genetic basis of flowering responses to seasonal cues', *Nature Reviews Genetics*. Nature Publishing Group, pp. 627–639. doi: 10.1038/nrg3291.

Anthony, R. G., James, P. E. and Jordan, B. R. (1996) 'Cauliflower (Brassica oleracea var. botrytis L.) curd development: the expression of meristem identity genes', *Journal of Experimental Botany*. Oxford Academic, 47(2), pp. 181–188. doi: 10.1093/JXB/47.2.181.

Arias, T. *et al.* (2014) 'Diversification times among Brassica (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence', *American Journal of Botany*. Hawkes, 101(1), pp. 86–91. doi: 10.3732/ajb.1300312.

Atwell, S. *et al.* (2010) 'Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines', *Nature*, 465(7298), pp. 627–631. doi: 10.1038/nature08800.

Aukerman, M. J. and Sakai, H. (2003) 'The Plant Cell Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes', *The Plant Cell*, 15, pp. 2730–2741. doi: 10.1105/tpc.016238.

Axelsson, T., Shavorskaya, O. and Lagercrantz, U. (2001) ' Multiple flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene ', *Genome*. Canadian Science Publishing, 44(5), pp. 856–864. doi: 10.1139/g01-082.

Babula, D. *et al.* (2003) 'Chromosomal mapping of Brassica oleracea based on ESTs from Arabidopsis thaliana: Complexity of the comparative map', *Molecular Genetics and Genomics*. Springer Verlag, 268(5), pp. 656–665. doi: 10.1007/s00438-002-0782-2.

Barker, G. C. *et al.* (2007) 'Novel Insights into Seed Fatty Acid Synthesis and Modification Pathways from Genetic Diversity and Quantitative Trait Loci Analysis of the Brassica C Genome 1', *Plant Physiology*. American Society, 144, pp. 1827–1842. doi: 10.1104/pp.107.096172.

Bäurle, I. and Dean, C. (2006) 'The Timing of Developmental Transitions in Plants', *Cell*. Cell Press, 125(4), pp. 655–664. doi: 10.1016/J.CELL.2006.05.005.

Beilstein, M. A. *et al.* (2010) 'Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(43), pp. 18724–18728. doi: 10.1073/pnas.0909766107.

Van Bel, M. *et al.* (2018) 'PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics', *Nucleic Acids Research*. Oxford Academic, 46(D1), pp. D1190–D1196. doi: 10.1093/nar/gkx1002.

Benjamini, Y. *et al.* (2005) *False Discovery Rate-Adjusted Multiple Confidence Intervals for Selected Para*, *Journal of the American Statistical Association*.

Birchler, J. A. and Veitia, R. A. (2010) 'The gene balance hypothesis: Implications for gene regulation, quantitative traits and evolution', *New Phytologist*. John Wiley & Sons, Ltd, pp. 54–62. doi: 10.1111/j.1469-8137.2009.03087.x.

Blazquez, M. A. *et al.* (1997) 'LEAFY expression and flower initiation in Arabidopsis', *Development*. The Company of Biologists, 124(19), pp. 3835–3844. doi: 10.1242/dev.124.19.3835.

Blümel, M., Dally, N. and Jung, C. (2015) 'Flowering time regulation in crops — what did we learn from Arabidopsis?', *Current Opinion in Biotechnology*. Elsevier Current Trends, 32, pp. 121–129. doi: 10.1016/J.COPBIO.2014.11.023.

Bohuon, E. J. R. *et al.* (1998) *The Association of Flowering Time Quantitative Trait Loci with Duplicated Regions and Candidate Loci in Brassica oleracea respondence of such candidate genes with QTL is com-may be several potential candidate genes within the*.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*. Oxford University Press, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Booij, R. and Struik, P. C. (1990) 'Effects of temperature on leaf and curd initiation in relation to juvenility of cauliflower', *Scientia Horticulturae*. Elsevier, 44(3–4), pp. 201–214. doi: 10.1016/0304-4238(90)90120-4.

Boss, P. K. *et al.* (no date) 'Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting'. doi: 10.1105/tpc.015958.

Bouché, F. *et al.* (2015) 'FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana', *Nucleic Acids Research*, 44, pp. 1167–1171. doi: 10.1093/nar/gkv1054.

Bowers, J. E. *et al.* (2003) 'Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events', *Nature*. Nature Publishing Group, 422(6930), pp. 433–438. doi: 10.1038/nature01521.

Bowman, J. L. *et al.* (1993) 'Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes', *Development*. The Company of Biologists, 119(3), pp. 721–743. doi: 10.1242/dev.119.3.721.

Bradbury, P. J. *et al.* (2007) 'TASSEL: software for association mapping of complex traits in diverse samples', *Bioinformatics*. Oxford Academic, 23(19), pp. 2633–2635. doi: 10.1093/bioinformatics/btm308.

Braynen, J. *et al.* (2021) 'Comparative transcriptome analysis revealed differential gene expression in multiple signaling pathways at flowering in polyploid Brassica rapa', *Cell and Bioscience*. BioMed Central, 11(1), pp. 1–13. doi: 10.1186/s13578-021-00528-1.

Brooks, L. *et al.* (2009) 'Microdissection of shoot meristem functional domains', *PLoS Genetics*. Public Library of Science, 5(5), p. e1000476. doi: 10.1371/journal.pgen.1000476.

Brown, L. (2021) ‘Horticulture Statistics 2020’, *DEFRA*, pp. 1–14.
Buhtz, A. *et al.* (2008) ‘Identification and characterization of small RNAs from the phloem of Brassica napus’, *Plant Journal*, 53(5), pp. 739–749. doi: 10.1111/j.1365-313X.2007.03368.x.
Busch, M. A., Bomblies, K. and Weigel, D. (1999) ‘Activation of a floral homeotic gene in Arabidopsis’, *Science*. Science, 285(5427), pp. 585–587. doi: 10.1126/science.285.5427.585.
Cai, G. *et al.* (2014) ‘A complex recombination pattern in the genome of allotetraploid Brassica napusas revealed by a high-density genetic map’, *PLoS ONE*. Public Library of Science, 9(10), p. e109910. doi: 10.1371/journal.pone.0109910.
Calderwood, A., Hepworth, J., *et al.* (2021) ‘Comparative transcriptomics reveals desynchronisation of gene expression during the floral transition between Arabidopsis and Brassica rapa cultivars’, *Quantitative Plant Biology*. Cambridge University Press, 2, pp. 1–13. doi: 10.1017/qpb.2021.6.
Calderwood, A., Lloyd, A., *et al.* (2021) ‘Total FLC transcript dynamics from divergent paralogue expression explains flowering diversity in Brassica napus’, *New Phytologist*. Blackwell Publishing Ltd, 229(6), pp. 3534–3548. doi: 10.1111/nph.17131.
Cartea, E. *et al.* (2019) ‘Seed Oil Quality of Brassica napus and Brassica rapa Germplasm from Northwestern Spain.’, *Foods (Basel, Switzerland)*. Multidisciplinary Digital Publishing Institute (MDPI), 8(8). doi: 10.3390/foods8080292.
Chalhoub, B. *et al.* (2014) ‘Erratum: Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome (Science (2014) 345:6199 (950-953))’, *Science*, p. 6199. doi: 10.1126/science.1260782.
Chan, T. E., Stumpf, M. P. H. and Babtie, A. C. (2017) ‘Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures’, *Cell Systems*. Cell Press, 5(3), pp. 251-267.e3. doi: 10.1016/j.cels.2017.08.014.
Cheng, J. Z. *et al.* (2017) ‘Research progress on the autonomous flowering time pathway in Arabidopsis’, *Physiology and Molecular Biology of Plants*. Springer India, pp. 477–485. doi: 10.1007/s12298-017-0458-3.
Chitwood, D. H., Maloof, J. N. and Sinha, N. R. (2013) ‘Dynamic transcriptomic profiles between tomato and a wild relative reflect distinct developmental architectures.’, *Plant physiology*. American Society of Plant Biologists, 162(2), pp. 537–52. doi: 10.1104/pp.112.213546.
Choi, K. *et al.* (2011) ‘The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors’, *Plant Cell*. American Society of Plant Biologists, 23(1), pp. 289–303. doi: 10.1105/tpc.110.075911.
Chouard, P. (1960) ‘Vernalization and its Relations to Dormancy’, *Annual Review of Plant Physiology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 11(1), pp. 191–238. doi: 10.1146/annurev.pp.11.060160.001203.
Clarke, J. H. and Dean, C. (1994) ‘Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana’, *MGG Molecular & General Genetics*. Springer-Verlag, 242(1), pp. 81–89. doi: 10.1007/BF00277351.
Clevenger, J. *et al.* (2015) ‘Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations’, *Molecular Plant*, pp. 831–846. doi: 10.1016/j.molp.2015.02.002.

Conant, G. C. and Wolfe, K. H. (2008) 'Turning a hobby into a job: How duplicated genes find new functions', *Nature Reviews Genetics*. Nature Publishing Group, pp. 938–950. doi: 10.1038/nrg2482.
Davis, S. J. (2009) 'Integrating hormones into the floral-transition pathway of Arabidopsis thaliana', *Plant, Cell and Environment*. Plant Cell Environ, 32(9), pp. 1201–1210. doi: 10.1111/j.1365-3040.2009.01968.x.
Duclos, D. V. and Björkman, T. (2008) 'Meristem identity gene expression during curd proliferation and flower initiation in Brassica oleracea', *Journal of Experimental Botany*. Oxford Academic, 59(2), pp. 421–433. doi: 10.1093/jxb/erm327.
Earl, D. A. and vonHoldt, B. M. (2012) 'STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method', *Conservation Genetics Resources*. Springer, 4(2), pp. 359–361. doi: 10.1007/s12686-011-9548-7.
Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*. Oxford Academic, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.
En Chai, L. *et al.* (2014) 'A review on the computational approaches for gene regulatory network construction', *Computers in Biology and Medicine*, 48, pp. 55–65. doi: 10.1016/j.compbiomed.2014.02.011.
Eriksson, S. *et al.* (2006) 'GA4 is the active gibberellin in the regulation of LEAFY transcription and Arabidopsis floral initiation', *Plant Cell*. American Society of Plant Biologists, 18(9), pp. 2172–2181. doi: 10.1105/tpc.106.042317.
Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software structure: a simulation study', *Molecular Ecology*. John Wiley & Sons, Ltd, 14(8), pp. 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x.
Flint-Garcia, S. A., Thornsberry, J. M. and Buckler, E. S. (2003) 'Structure of Linkage Disequilibrium in Plants', *Annual Review of Plant Biology*, 54(1), pp. 357–374. doi: 10.1146/annurev.arplant.54.031902.134907.
G, S. (1978) '"Estimating the Dimension of a Model."', *The Annals of Statistics*, 6(2), pp. 461–464. doi: 10.2307/2958889.
Gacek, K. *et al.* (2017) 'Genome-wide association study of genetic control of seed fatty acid biosynthesis in Brassica napus', *Frontiers in Plant Science*. Frontiers, 7, p. 2062. doi: 10.3389/fpls.2016.02062.
Gao, J. *et al.* (2014) 'Characterization of the floral transcriptome of Moso bamboo (Phyllostachys edulis) at different flowering developmental stages by transcriptome sequencing and RNA-seq analysis', *PLoS ONE*. Public Library of Science, 9(6), p. e98910. doi: 10.1371/journal.pone.0098910.
Gazave, E. *et al.* (2020) 'Genome-wide association study identifies acyl-lipid metabolism candidate genes involved in the genetic control of natural variation for seed fatty acid traits in Brassica napus L.', *Industrial Crops and Products*, 145, p. 112080. doi: 10.1016/j.indcrop.2019.112080.
Geraldo, N. *et al.* (2009) 'FRIGIDA delays flowering in Arabidopsis via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex', *Plant Physiology*. American Society of Plant Biologists, 150(3), pp. 1611–1618. doi: 10.1104/pp.109.137448.

Golicz, A. A. *et al.* (2016) 'The pangenome of an agronomically important crop plant Brassica oleracea', *Nature Communications*. Nature Publishing Group, 7, p. 13390. doi: 10.1038/ncomms13390.
Gorham, S. R. *et al.* (2018) 'HISTONE DEACETYLASE 19 and the flowering time gene FD maintain reproductive meristem identity in an age-dependent manner', *Journal of Experimental Botany*. Oxford University Press, 69(20), pp. 4757–4771. doi: 10.1093/jxb/ery239.
Goslin, K. *et al.* (2017) 'Transcription factor interplay between LEAFY and APETALA1/CAULIFLOWER during floral initiation', *Plant Physiology*, 174(2). doi: 10.1104/pp.17.00098.
Gregis, V. *et al.* (2008) 'AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis', *Plant Journal*. Plant J, 56(6), pp. 891–902. doi: 10.1111/j.1365-313X.2008.03648.x.
Gregis, V. *et al.* (2009) 'The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes', *Plant Journal*. John Wiley & Sons, Ltd, 60(4), pp. 626–637. doi: 10.1111/j.1365-313X.2009.03985.x.
Gregis, V. *et al.* (2013) 'Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis', *Genome Biology*. BioMed Central, 14(6), pp. 1–26. doi: 10.1186/gb-2013-14-6-r56.
Guo, C. *et al.* (2017) 'Repression of miR156 by miR159 Regulates the Timing of the Juvenile-to-Adult Transition in Arabidopsis', *The Plant Cell*. American Society of Plant Biologists, 29(6), p. tpc.00975.2016. doi: 10.1105/tpc.16.00975.
Guo, Y. *et al.* (2014) 'Mutations in single FT- and TFL1-paralogs of rapeseed (Brassica napus L.) and their impact on flowering time and yield components', *Frontiers in Plant Science*. Frontiers, 5(JUN), p. 282. doi: 10.3389/FPLS.2014.00282.
Gustafson-Brown, C., Savidge, B. and Yanofsky, M. F. (1994) 'Regulation of the arabidopsis floral homeotic gene APETALA1', *Cell*. Cell, 76(1), pp. 131–143. doi: 10.1016/0092-8674(94)90178-3.
Hand, D. J. and Atherton, J. G. (1987) 'Curd initiation in the cauliflower: I. Juvenility', *Journal of Experimental Botany*. Narnia, 38(12), pp. 2050–2058. doi: 10.1093/jxb/38.12.2050.
Harper, A. L. *et al.* (2012) 'Associative transcriptomics of traits in the polyploid crop species Brassica napus', *Nature Biotechnology*. Nature Publishing Group, 30(8), pp. 798–802. doi: 10.1038/nbt.2302.
Harper, A. L. *et al.* (2016) 'Molecular markers for tolerance of European ash (Fraxinus excelsior) to dieback disease identified using Associative Transcriptomics', *Scientific Reports*. Nature Publishing Group, 6(1), p. 19335. doi: 10.1038/srep19335.
Harrop, T. W. R. *et al.* (2016) 'Gene expression profiling of reproductive meristem types in early rice inflorescences by laser microdissection', *Plant Journal*. John Wiley & Sons, Ltd, 86(1), pp. 75–88. doi: 10.1111/tpj.13147.
Hasan, Y. *et al.* (2016) 'Quantitative trait loci controlling leaf appearance and curd initiation of cauliflower in relation to temperature', *Theoretical and Applied Genetics*, 129(7), pp. 1273–1288. doi: 10.1007/s00122-016-2702-6.
Havlickova, L. *et al.* (2018) 'Validation of an updated Associative Transcriptomics platform for the

polyploid crop species Brassica napus by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds', *Plant Journal*. John Wiley & Sons, Ltd (10.1111), 93(1), pp. 181–192. doi: 10.1111/tpj.13767.

He, Z. *et al.* (2015) 'Construction of Brassica A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research', *Data in Brief*. doi: 10.1016/j.dib.2015.06.016.

Hickman, R. *et al.* (2017) 'Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network.', *The Plant cell*. American Society of Plant Biologists, 29(9), pp. 2086–2105. doi: 10.1105/tpc.16.00958.

Hong, J. K. *et al.* (2011) 'Promoters of three Brassica rapa FLOWERING LOCUS C differentially regulate gene expression during growth and development in Arabidopsis', *Genes & Genomics*. The Genetics Society of Korea, 33(1), pp. 75–82. doi: 10.1007/s13258-010-0117-3.

Huala, E. *et al.* (2001) 'The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant', *Nucleic Acids Research*. Nucleic Acids Res, 29(1), pp. 102–105. doi: 10.1093/nar/29.1.102.

Huang, J. *et al.* (2014) 'Transcriptomic analysis of Asiatic lily in the process of vernalization via RNA-seq', *Molecular Biology Reports*, 41(6), pp. 3839–3852. doi: 10.1007/s11033-014-3250-2.

Huang, X. *et al.* (2012) 'Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm', *Nature Genetics*. Nature Publishing Group, 44(1), pp. 32–39. doi: 10.1038/ng.1018.

Huang, X. *et al.* (no date) 'Transcriptomic analysis of the regulation of stalk development in flowering Chinese cabbage (Brassica campestris) by RNA sequencing'. doi: 10.1038/s41598-017-15699-6.

Huijser, P. and Schmid, M. (2011) 'The control of developmental phase transitions in plants', *Development*. doi: 10.1242/dev.063511.

Hyun, K. gi, Noh, Y. S. and Song, J. J. (2017) 'Arabidopsis FRIGIDA stimulates EFS histone H3 Lys36 methyltransferase activity', *Plant Cell Reports*. Springer Verlag, 36(7), pp. 1183–1185. doi: 10.1007/s00299-017-2161-9.

I, D. O. *et al.* (2019) 'High ambient temperature leads to reduced FT expression and delayed flowering in Brassica rapa via a mechanism associated with H2A.Z dynamics', *The Plant journal : for cell and molecular biology*. Plant J, 100(2), pp. 343–356. doi: 10.1111/TPJ.14446.

Irwin, J. A. *et al.* (2012) 'Functional alleles of the flowering time regulator FRIGIDA in the Brassica oleracea genome', *BMC Plant Biology*. BioMed Central, 12(1), p. 21. doi: 10.1186/1471-2229-12-21.

Irwin, J. A. *et al.* (2016) 'Nucleotide polymorphism affecting FLC expression underpins heading date variation in horticultural brassicas', *The Plant journal : for cell and molecular biology*, 87(6), pp. 597–605. doi: 10.1111/tpj.13221.

Jacobsen, S. E. and Olszewski, N. E. (1993) 'Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction', *Plant Cell*. American Society of Plant Biologists, 5(8), pp. 887–896. doi: 10.1105/tpc.5.8.887.

Jaeger, K. E. *et al.* (2013) 'Interlocking feedback loops govern the dynamic behavior of the floral transition in Arabidopsis.', *The Plant cell*. American Society of Plant Biologists, 25(3), pp. 820–33.

doi: 10.1105/tpc.113.109355.
Jones, D. M. (2017) *Effects of gene multiplication on flowering time regulation in spring and winter varieties of Brassica napus*.
Jones, D. M. *et al.* (2017) 'Regulatory divergence of flowering time genes in the allopolyploid Brassica napus'. doi: 10.1101/178137.
Jones, D. M. *et al.* (2018) 'Spatio-temporal expression dynamics differ between homologues of flowering time genes in the allopolyploid Brassica napus', *Plant Journal*. Blackwell Publishing Ltd, 96(1), pp. 103–118. doi: 10.1111/tpj.14020.
Jung, C. and Müller, A. E. (2009) 'Flowering time control and applications in plant breeding', *Trends in Plant Science*. Elsevier Current Trends, pp. 563–573. doi: 10.1016/j.tplants.2009.07.005.
Jung, J.-H. *et al.* (no date) 'The GIGANTEA-Regulated MicroRNA172 Mediates Photoperiodic Flowering Independent of CONSTANS in Arabidopsis W OA'. doi: 10.1105/tpc.107.054528.
Juvenility, I., Hand, D. J. and Atherton, J. G. (1987) 'Curd Initiation in the Cauliflower', *Journal of Experimental Botany*, 38(197), pp. 2050–2058.
Kardailsky, I. *et al.* (1999) 'Activation tagging of the floral inducer FT', *Science*. American Association for the Advancement of Science, 286(5446), pp. 1962–1965. doi: 10.1126/science.286.5446.1962.
Kaul, S. *et al.* (2000) 'Analysis of the genome sequence of the flowering plant Arabidopsis thaliana', *Nature*. Nature Publishing Group, 408(6814), pp. 796–815. doi: 10.1038/35048692.
Kearse, M. *et al.* (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*. Oxford Academic, 28(12), pp. 1647–1649. doi: 10.1093/bioinformatics/bts199.
Kendon, M. *et al.* (2021) 'State of the UK Climate 2020', *International Journal of Climatology*. John Wiley & Sons, Ltd, 41(S2), pp. 1–76. doi: 10.1002/joc.7285.
Kim, D., Langmead, B. and Salzberg, S. L. (2015) 'HISAT: a fast spliced aligner with low memory requirements', *Nature Methods 2015 12:4*. Nature Publishing Group, 12(4), pp. 357–360. doi: 10.1038/nmeth.3317.
Kim, K. H. *et al.* (2021) 'Rna-seq analysis of gene expression changes related to delay of flowering time under drought stress in tropical maize', *Applied Sciences (Switzerland)*. Multidisciplinary Digital Publishing Institute, 11(9), p. 4273. doi: 10.3390/app11094273.
Kim, S. Y. *et al.* (2005) 'Establishment of the vernalization-responsive, winter-annual habit in Arabidopsis requires a putative histone H3 methyl transferase', *Plant Cell*. American Society of Plant Biologists, 17(12), pp. 3301–3310. doi: 10.1105/tpc.105.034645.
Klepikova, A. V. *et al.* (2015) 'RNA-seq analysis of an apical meristem time series reveals a critical point in Arabidopsis thaliana flower initiation', *BMC Genomics*. BioMed Central, 16(1), p. 466. doi: 10.1186/s12864-015-1688-9.
Klepikova, A. V. *et al.* (2016) 'A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling', *Plant Journal*. John Wiley & Sons, Ltd (10.1111), 88(6), pp. 1058–1070. doi: 10.1111/tpj.13312.
Kobayashi, Y. *et al.* (1999) 'A pair of related genes with antagonistic roles in mediating flowering

signals', *Science*. American Association for the Advancement of Science, 286(5446), pp. 1960–1962. doi: 10.1126/science.286.5446.1960.
Kohonen, T. (1997) 'Exploration of very large databases by self-organizing maps', *IEEE International Conference on Neural Networks - Conference Proceedings*, 1. doi: 10.1109/ICNN.1997.611622.
Koornneeff, M., Dellaert, L. W. M. and van der Veen, J. H. (1982) 'EMS- and relation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh', *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. Elsevier, 93(1), pp. 109–123. doi: 10.1016/0027-5107(82)90129-4.
Kop, E. *et al.* (2003) 'Genetic analysis of the bracting trait in cauliflower and broccoli Characterisation of Genetic Variation in Chemical Nutritional Composition of bambara groundnut (Vigna subterranea) View project Genetic analysis of the bracting trait in cauliflower and broccoli'. doi: 10.1016/S0168-9452(03)00068-2.
Kullback, S. and Leibler, R. A. (1951) 'On Information and Sufficiency', *The Annals of Mathematical Statistics*, 22(1), pp. 79–86. doi: 10.1214/aoms/1177729694.
Labana, K. S. and Gupta, M. L. (1993) 'Importance and Origin', in. Springer, Berlin, Heidelberg, pp. 1–7. doi: 10.1007/978-3-662-06166-4_1.
Lagercrantz, U. and Lydiate, D. J. (1996) 'Comparative genome mapping in Brassica', *Genetics*. Oxford University Press, 144(4), pp. 1903–1910. doi: 10.1093/genetics/144.4.1903.
Lan, T. H. *et al.* (2000) 'An EST-enriched comparative map of Brassica oleracea and Arabidopsis thaliana', *Genome Research*. Genome Res, 10(6), pp. 776–788. doi: 10.1101/gr.10.6.776.
Laufs, P. *et al.* (1998) 'Cellular parameters of the shoot apical meristem in Arabidopsis', *Plant Cell*. Oxford Academic, 10(8), pp. 1375–1389. doi: 10.1105/tpc.10.8.1375.
Lauri, P. E. and Normand, F. (2017) 'Are leaves only involved in flowering? Bridging the gap between structural botany and functional morphology', *Tree Physiology*. Oxford Academic, pp. 1137–1139. doi: 10.1093/treephys/tpx068.
Lee, J. *et al.* (2008) 'SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY', *Plant Journal*. Plant J, 55(5), pp. 832–843. doi: 10.1111/j.1365-313X.2008.03552.x.
Lee, J. and Lee, I. (2010) 'Regulation and function of SOC1, a flowering pathway integrator', *Journal of Experimental Botany*. J Exp Bot, pp. 2247–2254. doi: 10.1093/jxb/erq098.
Leijten, W. *et al.* (2018) 'Translating flowering time from arabidopsis thaliana to brassicaceae and asteraceae crop species', *Plants*. Multidisciplinary Digital Publishing Institute, p. 111. doi: 10.3390/plants7040111.
Li, D. *et al.* (2008) 'A Repressor Complex Governs the Integration of Flowering Signals in Arabidopsis', *Developmental Cell*. Cell Press, 15(1), pp. 110–120. doi: 10.1016/j.devcel.2008.05.002.
Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*. Oxford Academic, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.
Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform.', *Bioinformatics (Oxford, England)*. Oxford University Press, 25(14), pp. 1754–60. doi: 10.1093/bioinformatics/btp324.

Li, Y., Wu, F.-X. and Ngom, A. (2018) ‘A review on machine learning principles for multi-view biological data integration’, *Briefings in Bioinformatics*. Oxford Academic, 19(2), pp. 325–340. doi: 10.1093/BIB/BBW113.

Liljegren, S. J. *et al.* (1999) ‘Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate’, *Plant Cell*. Plant Cell, 11(6), pp. 1007–1018. doi: 10.1105/tpc.11.6.1007.

Lin, S.-I. *et al.* (2005) ‘Differential Regulation of FLOWERING LOCUS C Expression by Vernalization in Cabbage and Arabidopsis 1’, *Plant physiology*, 137, pp. 1037–1048. doi: 10.1104/pp.104.058974.

Liu, C. *et al.* (2008) ‘Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis’, *Development*. The Company of Biologists, 135(8), pp. 1481–1491. doi: 10.1242/dev.020255.

Liu, C. *et al.* (2009) ‘Regulation of Floral Patterning by Flowering Time Genes’, *Developmental Cell*. Cell Press, 16(5), pp. 711–722. doi: 10.1016/j.devcel.2009.03.011.

Liu, S. *et al.* (2014) ‘The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes’, *Nature Communications*, 5(1), pp. 1–11. doi: 10.1038/ncomms4930.

Lowman, A. and Purugganan, M. (1999) ‘Duplication of the Brassica oleracea APETALA1floral homeotic gene and the evolution of domesticated cauliflower’, *Journal of Heredity*. Oxford Academic, 90(5), pp. 514–520. doi: 10.1093/JHERED/90.5.514.

Lu, X. *et al.* (2017) ‘RNA-seq analysis of apical meristem reveals integrative regulatory network of ROS and chilling potentially related to flowering in Litchi chinensis’, *Scientific Reports*. Nature Publishing Group, 7(1), p. 10183. doi: 10.1038/s41598-017-10742-y.

Lukens, L. *et al.* (2003) ‘Comparison of a Brassica oleracea genetic map with the genome of Arabidopsis thaliana’, *Genetics*. Oxford University Press, 164(1), pp. 359–372. doi: 10.1093/genetics/164.1.359.

Maggioni, L. *et al.* (2010) ‘Origin and domestication of cole crops (Brassica oleracea L.): Linguistic and literary considerations’, *Economic Botany*. Springer, 64(2), pp. 109–123. doi: 10.1007/s12231-010-9115-2.

Maggioni, L. (2015) ‘Domestication of Brassica oleracea L.’, *Doctoral Thesis*, pp. 1–110.

Mandel, M. A. and Yanofsky, M. F. (1995) ‘A gene triggering flower formation in Arabidopsis’, *Nature*. Nature Publishing Group, pp. 522–524. doi: 10.1038/377522a0.

Marbach, D. *et al.* (2012) ‘Wisdom of crowds for robust gene network inference’, *Nature Methods*. Nature Publishing Group, 9(8), pp. 796–804. doi: 10.1038/nmeth.2016.

Matschegewski, C. *et al.* (2015) ‘Genetic variation of temperature-regulated curd induction in cauliflower: Elucidation of floral transition by genome-wide association mapping and gene expression analysis’, *Frontiers in Plant Science*. Frontiers Research Foundation, 6(September), p. 720. doi: 10.3389/fpls.2015.00720.

McClung, C. R. (2006) ‘Plant circadian rhythms’, *Plant Cell*. American Society of Plant Biologists, pp. 792–803. doi: 10.1105/tpc.106.040980.

Meier, U. (2001) *Growth stages of mono-and dicotyledonous plants BBCH Monograph Federal Biological Research Centre for Agriculture and Forestry*.

Mercatelli, D. *et al.* (2020) 'Gene regulatory network inference resources: A practical overview', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. Elsevier, p. 194430. doi: 10.1016/j.bbagrm.2019.194430.
Merchant, N. *et al.* (2016) 'The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences', *PLoS Biology*. Public Library of Science, 14(1), p. e1002342. doi: 10.1371/journal.pbio.1002342.
Michaels, S. D. *et al.* (2003) 'AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization', *The Plant Journal*. John Wiley & Sons, Ltd, 33(5), pp. 867–874. doi: 10.1046/J.1365-313X.2003.01671.X.
Michaels, S. D. and Amasino, R. M. (2001) 'Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGADA and autonomous pathway mutations but not responsiveness to vernalization', *Plant Cell*. Plant Cell, 13(4), pp. 935–941. doi: 10.1105/tpc.13.4.935.
Mizukami, Y. and Ma, H. (1997) 'Determination of arabidopsis floral meristem identity by AGAMOUS', *Plant Cell*. American Society of Plant Physiologists, 9(3), pp. 393–408. doi: 10.1105/tpc.9.3.393.
Moon, J. *et al.* (2003) 'The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis', *Plant Journal*. John Wiley & Sons, Ltd, 35(5), pp. 613–623. doi: 10.1046/j.1365-313X.2003.01833.x.
Nagaharu, U. (1935) 'Genome-analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization', *Japanese Journal of Botany*, 7, pp. 389–452.
Napp-Zinn, K. (1961) 'Über die Bedeutung genetischer Untersuchungen an kältebedürftigen Pflanzen für die Aufklärung von Vernalisationserscheinungen', *Der Züchter*. Springer-Verlag, 31(4), pp. 128–135. doi: 10.1007/BF00709087.
Neath, A. A. and Cavanaugh, J. E. (2012) 'The Bayesian information criterion: Background, derivation, and applications', *Wiley Interdisciplinary Reviews: Computational Statistics*. John Wiley & Sons, Ltd, 4(2), pp. 199–203. doi: 10.1002/wics.199.
O'Neill, C. M. and Bancroft, I. (2000) 'Comparative physical mapping of segments of the genome of Brassica oleracea var. alboglabra that are homoeologous to sequenced regions of chromosomes 4 and 5 of Arabidopsis thaliana', *Plant Journal*. Plant J, 23(2), pp. 233–243. doi: 10.1046/j.1365-313X.2000.00781.x.
Okazaki, K. *et al.* (2007) 'Mapping and characterization of FLC homologs and QTL analysis of flowering time in Brassica oleracea', *Theoretical and Applied Genetics*. Theor Appl Genet, 114(4), pp. 595–608. doi: 10.1007/s00122-006-0460-6.
Otto, S. P. (2007) 'The Evolutionary Consequences of Polyploidy', *Cell*. Elsevier B.V., pp. 452–462. doi: 10.1016/j.cell.2007.10.022.
Papp, B., Pál, C. and Hurst, L. D. (2003) 'Dosage sensitivity and the evolution of gene families in yeast', *Nature*. Nature Publishing Group, 424(6945), pp. 194–197. doi: 10.1038/nature01771.
Parcy, F., Bomblies, K. and Weigel, D. (2002) 'Interaction of *LEAFY* , *AGAMOUS* and *TERMINAL FLOWER1* in maintaining floral meristem identity in *Arabidopsis*', *Development*. The Company of

Biologists, 129(10), pp. 2519–2527. doi: 10.1242/dev.129.10.2519.
Park, J. Y. *et al.* (2005) 'Physical mapping and microsynteny of Brassica rapa ssp. pekinensis genome corresponding to a 222 kbp gene-rich region of Arabidopsis chromosome 4 and partially duplicated on chromosome 5', *Molecular Genetics and Genomics*. Mol Genet Genomics, 274(6), pp. 579–588. doi: 10.1007/s00438-005-0041-4.
Parkin, I. A. *et al.* (2014) 'Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea', *Genome Biology*. BioMed Central, 15(6), p. R77. doi: 10.1186/gb-2014-15-6-r77.
Penfold, C. A. *et al.* (2015) 'CSI: A Nonparametric Bayesian Approach to Network Inference from Multiple Perturbed Time Series Gene Expression Data', *Statistical Applications in Genetics and Molecular Biology*, 14(3), pp. 307–310.
Penfold, C. A. and Buchanan-Wollaston, V. (2014) 'Modelling transcriptional networks in leaf senescence', *Journal of experimental botany*, 65(14), pp. 3859–3873. doi: 10.1093/jxb/eru054.
Penfold, C. A. and Wild, D. L. (2011) 'How to infer gene networks from expression profiles, revisited.', *Interface focus*. Royal Society, 1(6), pp. 857–70. doi: 10.1098/rsfs.2011.0053.
Pertea, M. *et al.* (2015) 'StringTie enables improved reconstruction of a transcriptome from RNA-seq reads', *Nature Biotechnology*. Nature Publishing Group, 33(3), pp. 290–295. doi: 10.1038/nbt.3122.
Pires, J. C. *et al.* (2004) 'Flowering time divergence and genomic rearrangements in resynthesized Brassica polyploids (Brassicaceae)', in *Biological Journal of the Linnean Society*. Oxford Academic, pp. 675–688. doi: 10.1111/j.1095-8312.2004.00350.x.
Poethig, R. S. (2009) 'Small RNAs and developmental timing in plants', *Current Opinion in Genetics & Development*. Elsevier Current Trends, 19(4), pp. 374–378. doi: 10.1016/J.GDE.2009.06.001.
Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) 'Inference of population structure using multilocus genotype data.', *Genetics*, 155(2), pp. 945–59.
Prom, L. K. *et al.* (2019) 'GWAS analysis of sorghum association panel lines identifies SNPs associated with disease response to Texas isolates of Colletotrichum sublineola', *Theoretical and Applied Genetics*, 132, pp. 1389–1396. doi: 10.1007/s00122-019-03285-5.
Putterill, J. *et al.* (1995) 'The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors', *Cell*. Cell, 80(6), pp. 847–857. doi: 10.1016/0092-8674(95)90288-0.
Putterill, J., Laurie, R. and Macknight, R. (2004) 'It's time to flower: The genetic control of flowering time', *BioEssays*. Wiley Subscription Services, Inc., A Wiley Company, pp. 363–373. doi: 10.1002/bies.20021.
Rafalski, J. A. (2010) 'Association genetics in crop improvement', *Current Opinion in Plant Biology*. Elsevier Current Trends, 13(2), pp. 174–180. doi: 10.1016/J.PBI.2009.12.004.
Raman, H. *et al.* (2019) 'GWAS hints at pleiotropic roles for FLOWERING LOCUS T in flowering time and yield-related traits in canola', *BMC Genomics*, 20(1). doi: 10.1186/s12864-019-5964-y.
Rana, D. *et al.* (2004) 'Conservation of the microstructure of genome segments in Brassica napus and its diploid relatives', *Plant Journal*. Plant J, 40(5), pp. 725–733. doi: 10.1111/j.1365-313X.2004.02244.x.

Ratcliffe, O. J., Bradley, D. J. and Coen, E. S. (1999) 'Separation of shoot and floral identity in Arabidopsis', *Development*. Schultz and Haughn, 126(6), pp. 1109–1120. doi: 10.1242/dev.126.6.1109.

Razi, H. *et al.* (2008) 'Does sequence polymorphism of FLC paralogues underlie flowering time QTL in Brassica oleracea?', *Theoretical and Applied Genetics*, 116(2), pp. 179–192. doi: 10.1007/s00122-007-0657-3.

Ridge, S. *et al.* (2015) 'The role of BoFLC2 in cauliflower (Brassica oleracea var. botrytis L.) reproductive development', *Journal of Experimental Botany*. Oxford Academic, 66(1), pp. 125–135. doi: 10.1093/jxb/eru408.

Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer.', *Nature biotechnology*. NIH Public Access, 29(1), pp. 24–6. doi: 10.1038/nbt.1754.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*. Oxford Academic, 26(1), pp. 139–140. doi: 10.1093/bioinformatics/btp616.

Romera-Branchat, M., Andrés, F. and Coupland, G. (2014) 'Flowering responses to seasonal cues: What's new?', *Current Opinion in Plant Biology*. doi: 10.1016/j.pbi.2014.07.006.

Romero Navarro, J. A. *et al.* (2017) 'A study of allelic diversity underlying flowering-time adaptation in maize landraces', *Nature Genetics*, 49(3), pp. 476–480. doi: 10.1038/ng.3784.

Rosen, A. *et al.* (2018) 'Genome-Based Prediction of Time to Curd Induction in Cauliflower', *Frontiers in Plant Science*. Frontiers, 9, p. 78. doi: 10.3389/fpls.2018.00078.

Ruiz-García, L. *et al.* (1997) 'Different roles of flowering-time genes in the activation of floral initiation genes in Arabidopsis', *Plant Cell*. American Society of Plant Physiologists, 9(11), pp. 1921–1934. doi: 10.1105/tpc.9.11.1921.

Saint-Antoine, M. M. and Singh, A. (2020) 'Network inference in systems biology: recent developments, challenges, and applications', *Current Opinion in Biotechnology*. Elsevier Current Trends, 63, pp. 89–98. doi: 10.1016/J.COPBIO.2019.12.002.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees.', *Molecular biology and evolution*. Oxford Academic, 4(4), pp. 406–425. doi: 10.1093/oxfordjournals.molbev.a040454.

Salter, P. J. and James, J. M. (1974) 'Further Studies on the Effects of Cold Treatment of Transplants on Crop Maturity Characteristics of Cauliflower', *Journal of Horticultural Science*. Taylor & Francis, 49(4), pp. 329–342. doi: 10.1080/00221589.1974.11514587.

Samach, A. *et al.* (2000) 'Distinct roles of constans target genes in reproductive development of Arabidopsis', *Science*. Science, 288(5471), pp. 1613–1616. doi: 10.1126/science.288.5471.1613.

Schiessl, S. V. *et al.* (2017) 'Flowering time gene variation in Brassica species shows evolutionary principles', *Frontiers in Plant Science*. Frontiers, 8, p. 1742. doi: 10.3389/fpls.2017.01742.

Schönrock, N. *et al.* (2006) 'Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway', *Genes and Development*. Cold Spring Harbor Laboratory Press, 20(12), pp. 1667–1678. doi: 10.1101/gad.377206.

Schranz, M. E. *et al.* (2002) 'Characterization and effects of the replicated flowering time gene FLC

in Brassica rapa', *Genetics*, 162(3), pp. 1457–1468. doi: 10.1093/genetics/162.3.1457.
Scotr Poethig, R. (1990) 'Phase Change and the Regulation of Shoot Morphogenesis in Plants', *Science*, 250(4983), pp. 923–930.
Searle, I. *et al.* (2006) 'The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis', *Genes and Development*. Genes Dev, 20(7), pp. 898–912. doi: 10.1101/gad.373506.
Serrano-Mislata, A. *et al.* (2017) 'Regulatory interplay between LEAFY, APETALA1/CAULIFLOWER and TERMINAL FLOWER1: New insights into an old relationship', *Plant Signaling and Behavior*. Taylor & Francis, 12(10), p. e1370164. doi: 10.1080/15592324.2017.1370164.
Shah, S. *et al.* (2018) 'Whole-transcriptome analysis reveals genetic factors underlying flowering time regulation in rapeseed (Brassica napus L.)', *Plant Cell and Environment*. John Wiley & Sons, Ltd, 41(8), pp. 1935–1947. doi: 10.1111/pce.13353.
Shannon, P. *et al.* (2003) 'Cytoscape: A software Environment for integrated models of biomolecular interaction networks', *Genome Research*. Cold Spring Harbor Laboratory Press, 13(11), pp. 2498–2504. doi: 10.1101/gr.1239303.
Shannon, S. and Meeks-Wagner, D. R. (1991) 'A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development.', *The Plant Cell*. Oxford University Press (OUP), 3(9), pp. 877–892. doi: 10.1105/tpc.3.9.877.
Sheldon, C. C. *et al.* (1999) 'The FLF MADS box gene: A repressor of flowering in Arabidopsis regulated by vernalization and methylation', *Plant Cell*, 11(3), pp. 445–458. doi: 10.1105/tpc.11.3.445.
Shivaraj, S. *et al.* (2014) 'Natural genetic variation in MIR172 isolated from Brassica species', *BIOLOGIA PLANTARUM*, 58(4), pp. 627–640. doi: 10.1007/s10535-014-0441-6.
Simpson, G. G. (2004) 'The autonomous pathway: Epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time', *Current Opinion in Plant Biology*. Curr Opin Plant Biol, pp. 570–574. doi: 10.1016/j.pbi.2004.07.002.
Simpson, G. G. and Dean, C. (2002) 'Arabidopsis, the Rosetta stone of flowering time?', *Science*. American Association for the Advancement of Science, pp. 285–289. doi: 10.1126/science.296.5566.285.
Smith, L. B. and King, G. J. (2000) 'The distribution of BoCAL-a alleles in Brassica oleracea is consistent with a genetic model for curd development and domestication of the cauliflower', *Molecular Breeding 2000 6:6*. Springer, 6(6), pp. 603–613. doi: 10.1023/A:1011370525688.
Song, J., Irwin, J. and Dean, C. (2013) 'Remembering the prolonged cold of winter', *Current Biology*, pp. R807–R811. doi: 10.1016/j.cub.2013.07.027.
Song, L., Langfelder, P. and Horvath, S. (2012) 'Comparison of co-expression measures: mutual information, correlation, and model based indices', *BMC Bioinformatics 2012 13:1*. BioMed Central, 13(1), pp. 1–21. doi: 10.1186/1471-2105-13-328.
Spanudakis, E. and Jackson, S. (2014) 'The role of microRNAs in the control of flowering time', *Journal of Experimental Botany*. Oxford Academic, 65(2), pp. 365–380. doi: 10.1093/jxb/ert453.

Srikanth, A. and Schmid, M. (2011) 'Regulation of flowering time: All roads lead to Rome', *Cellular and Molecular Life Sciences*. Springer, pp. 2013–2037. doi: 10.1007/s00018-011-0673-y.
Storey, J. *et al.* (2019) *Qvalue*, *qvalue: Q-value estimation for false discovery rate control. R package version 2.18.0*. Available at: http://github.com/jdstorey/qvalue (Accessed: 20 March 2020).
Suárez-López, P. *et al.* (2001) 'CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis', *Nature*. Nature Publishing Group, 410(6832), pp. 1116–1120. doi: 10.1038/35074138.
Sunkar, R. and Jagadeeswaran, G. (2008) 'In silico identification of conserved microRNAs in large number of diverse plant species', *BMC Plant Biology*. BioMed Central, 8(1), p. 37. doi: 10.1186/1471-2229-8-37.
Tallis, A. (2011) *Investigating Vernalization in Brassica oleracea*.
Team, R. C. (2013) 'The R Project for Statistical Computing', *Http://Www.R-Project.Org/*, pp. 1–12.
Teotia, S. and Tang, G. (2015) 'To bloom or not to bloom: Role of micrornas in plant flowering', *Molecular Plant*, pp. 359–377. doi: 10.1016/j.molp.2014.12.018.
Thornley, J. H. M. (1972) 'A model of a biochemical switch, and its application to flower initiation', *Annals of Botany*. Oxford University Press, 36(5), pp. 861–871. doi: 10.1093/oxfordjournals.aob.a084648.
Town, C. D. *et al.* (2006) 'Comparative genomics of Brassica oleracea and Arabidopsis thaliana reveal gene loss, fragmentation, and dispersal after polyploidy', *Plant Cell*. Oxford University Press, 18(6), pp. 1348–1359. doi: 10.1105/tpc.106.041665.
Trick, M. *et al.* (2009) 'Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing', *Plant Biotechnology Journal*, 7(4), pp. 334–346. doi: 10.1111/j.1467-7652.2008.00396.x.
Tyagi, S. *et al.* (2018) 'Natural variation in Brassica FT homeologs influences multiple agronomic traits including flowering time, silique shape, oil profile, stomatal morphology and plant height in B. juncea', *Plant Science*. Elsevier, 277, pp. 251–266. doi: 10.1016/J.PLANTSCI.2018.09.018.
University of Warwick (2009) *Developing tools for growers and breeders to enable the predictable manipulation of flowering*.
Valentim, F. L. *et al.* (2015) 'A quantitative and dynamic model of the arabidopsis flowering time gene regulatory network', *PLoS ONE*. Public Library of Science, 10(2), p. e0116973. doi: 10.1371/journal.pone.0116973.
Walley, P. G. *et al.* (2012) 'Developing genetic resources for pre-breedingin Brassica oleracea L.: An overview of the UK perspective', *Journal of Plant Biotechnology*, pp. 62–68. doi: 10.5010/JPB.2012.39.1.062.
Wang, J. W., Czech, B. and Weigel, D. (2009) 'miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana', *Cell*, 138(4), pp. 738–749. doi: 10.1016/j.cell.2009.06.014.
Wehrens, R. and Buydens, L. M. C. (2007) 'Self-and Super-organizing Maps in R: The kohonen Package', *JSS Journal of Statistical Software*, 21.
Weigel, D. *et al.* (1992) 'LEAFY controls floral meristem identity in Arabidopsis', *Cell*. Cell Press,

69(5), pp. 843–859. doi: 10.1016/0092-8674(92)90295-N.
Wigge, P. A. *et al.* (2005) 'Integration of spatial and temporal information during floral induction in Arabidopsis', *Science*. American Association for the Advancement of Science, 309(5737), pp. 1056–1059. doi: 10.1126/science.1114358.
Willmann, M. R. and Poethig, R. S. (2005) 'Time to grow up: the temporal role of smallRNAs in plants.', *Current opinion in plant biology*. NIH Public Access, 8(5), pp. 548–52. doi: 10.1016/j.pbi.2005.07.008.
Wilson, R. N., Heckman, J. W. and Somerville, C. R. (1992) 'Gibberellin is required for flowering in Arabidopsis thaliana under short days', *Plant Physiology*. American Society of Plant Biologists, 100(1), pp. 403–408. doi: 10.1104/pp.100.1.403.
Wu, G. *et al.* (2009) 'The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis', *Cell*. Cell Press, 138(4), pp. 750–759. doi: 10.1016/J.CELL.2009.06.031.
Wu, G. and Poethig, R. S. (2006) 'Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3.', *Development (Cambridge, England)*. NIH Public Access, 133(18), pp. 3539–47. doi: 10.1242/dev.02521.
Wurr, D. C. E. *et al.* (2016) 'A model of cauliflower curd growth to predict when curds reach a specified size', *Journal of Horticultural Science*. Taylor & Francis, 65(5), pp. 555–564. doi: 10.1080/00221589.1990.11516093.
Xanthopoulou, A. *et al.* (2021) 'A comprehensive RNA-Seq-based gene expression atlas of the summer squash (Cucurbita pepo) provides insights into fruit morphology and ripening mechanisms', *BMC Genomics*. BioMed Central, 22(1), pp. 1–17. doi: 10.1186/s12864-021-07683-2.
Xu, L. *et al.* (2015) 'Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (Brassica napus L.)', *DNA Research*. Oxford Academic, 23(1), pp. 43–52. doi: 10.1093/dnares/dsv035.
Xu, M. *et al.* (2016) 'Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana', *PLoS Genetics*. doi: 10.1371/journal.pgen.1006263.
Yadav, R. K. *et al.* (2009) 'Gene expression map of the Arabidopsis shoot apical mer i stem stem cell n i che', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(12), pp. 4941–4946. doi: 10.1073/pnas.0900843106.
Yang, L. *et al.* (2013) 'Sugar promotes vegetative phase change in Arabidopsis thaliana by repressing the expression of MIR156A and MIR156C', *eLife*. eLife Sciences Publications Limited, 2, p. e00260. doi: 10.7554/eLife.00260.
Yang, Y. W. *et al.* (1999) 'Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages', *Journal of Molecular Evolution*. J Mol Evol, 48(5), pp. 597–604. doi: 10.1007/PL00006502.
Yoo, S. K. *et al.* (2005) 'CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis', *Plant Physiology*. Oxford University Press, 139(2), pp. 770–778. doi: 10.1104/pp.105.066928.
Yu, G. *et al.* (2017) 'ggtree: an r package for visualization and annotation of phylogenetic trees with

their covariates and other associated data', *Methods in Ecology and Evolution*. British Ecological Society, 8(1), pp. 28–36. doi: 10.1111/2041-210X.12628.
Yu, H. *et al.* (2004) 'Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development', *Nature Genetics*. Nature Publishing Group, 36(2), pp. 157–161. doi: 10.1038/ng1286.
Yu, S. *et al.* (2013) 'Sugar is an endogenous cue for juvenile-to-adult phase transition in plants', *eLife*. eLife Sciences Publications Limited, 2, p. e00269. doi: 10.7554/eLife.00269.
Zeevaart, J. A. D. (1976) 'Physiology of Flower Formation', *Annual Review of Plant Physiology*. Annual Reviews, 27(1), pp. 321–348. doi: 10.1146/annurev.pp.27.060176.001541.
Zhao, M. *et al.* (2021) 'A comprehensive overview and critical evaluation of gene regulatory network inference technologies', *Briefings in Bioinformatics*. Oxford University Press (OUP), 2021(00), pp. 1–15. doi: 10.1093/BIB/BBAB009.
Zhou, B. *et al.* (2012) 'Hydrogen peroxide and nitric oxide promote reproductive growth in Litchi chinensis', *Biologia Plantarum*. Springer Netherlands, 56(2), pp. 321–329. doi: 10.1007/s10535-012-0093-3.
Zhu, Q.-H. and Helliwell, C. A. (2011) 'Regulation of flowering time and floral patterning by miR172', *Journal of Experimental Botany*, 62(2), pp. 487–495. doi: 10.1093/jxb/erq295.
Ziolkowski, P. A. *et al.* (2006) 'Genome evolution in Arabidopsis/Brassica: Conservation and divergence of ancient rearranged segments and their breakpoints', *Plant Journal*. John Wiley & Sons, Ltd, 47(1), pp. 63–74. doi: 10.1111/j.1365-313X.2006.02762.x.

Appendix A

Supplementary Figures and Tables for Chapter 2.

S.1A: Details of phenotyped panel, with associated crop type and subspecies information.

Genotype	Crop Type	Subspecies	Genetic Status
DJ3290	Brussels Sprouts	*Brassica oleracea var. gemmifera*	DH
GK040099	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT050381	Kale	*Brassica oleracea var. alboglabra*	DH
GT060867	Kale	*Brassica oleracea var. alboglabra*	DH
GT060871	Broccoli	*Brassica oleracea var. italica*	DH
GT061882	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT070804	Cabbage	*Brassica oleracea var. costata DC*	DH
GT080341	Broccoli	*Brassica oleracea var. italica*	DH
GT080486	Kale	*Brassica oleracea var. alboglabra*	DH
GT080693	Broccoli	*Brassica oleracea var. italica*	DH
GT080713	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT080760	Kale	*Brassica oleracea var. viridis*	DH
GT080767	Kale	*Brassica oleracea var. alboglabra*	DH
GT080843	Broccoli	*Brassica oleracea var. italica*	DH
GT080849	Broccoli	*Brassica oleracea var. italica*	DH
GT080869	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT080876	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT080891	Kale	*Brassica oleracea var. sabellica*	DH
GT081012	Kale	*Brassica oleracea var. sabellica*	DH
GT081062	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT081103	Broccoli	*Brassica oleracea var. italica*	DH
GT081140	Broccoli	*Brassica oleracea var. italica*	DH
GT081150	Broccoli	*Brassica oleracea var. italica*	DH
GT081395	Broccoli	*Brassica oleracea var. italica*	DH
GT090058	Cabbage	*Brassica oleracea var. capitata f. alba*	DH
GT090341	Cabbage	*Brassica oleracea var. capitata f. alba*	DH
GT100062	Broccoli	*Brassica oleracea var. italica*	DH
GT100065	Broccoli	*Brassica oleracea var. italica*	DH
GT100067	Kale	*Brassica oleracea var. alboglabra*	S4
GT100120	Broccoli	*Brassica oleracea var. italica*	DH
GT110206	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT110221	Kale	*Brassica oleracea var. alboglabra*	DH
GT110244	Broccoli	*Brassica oleracea var. italica*	DH

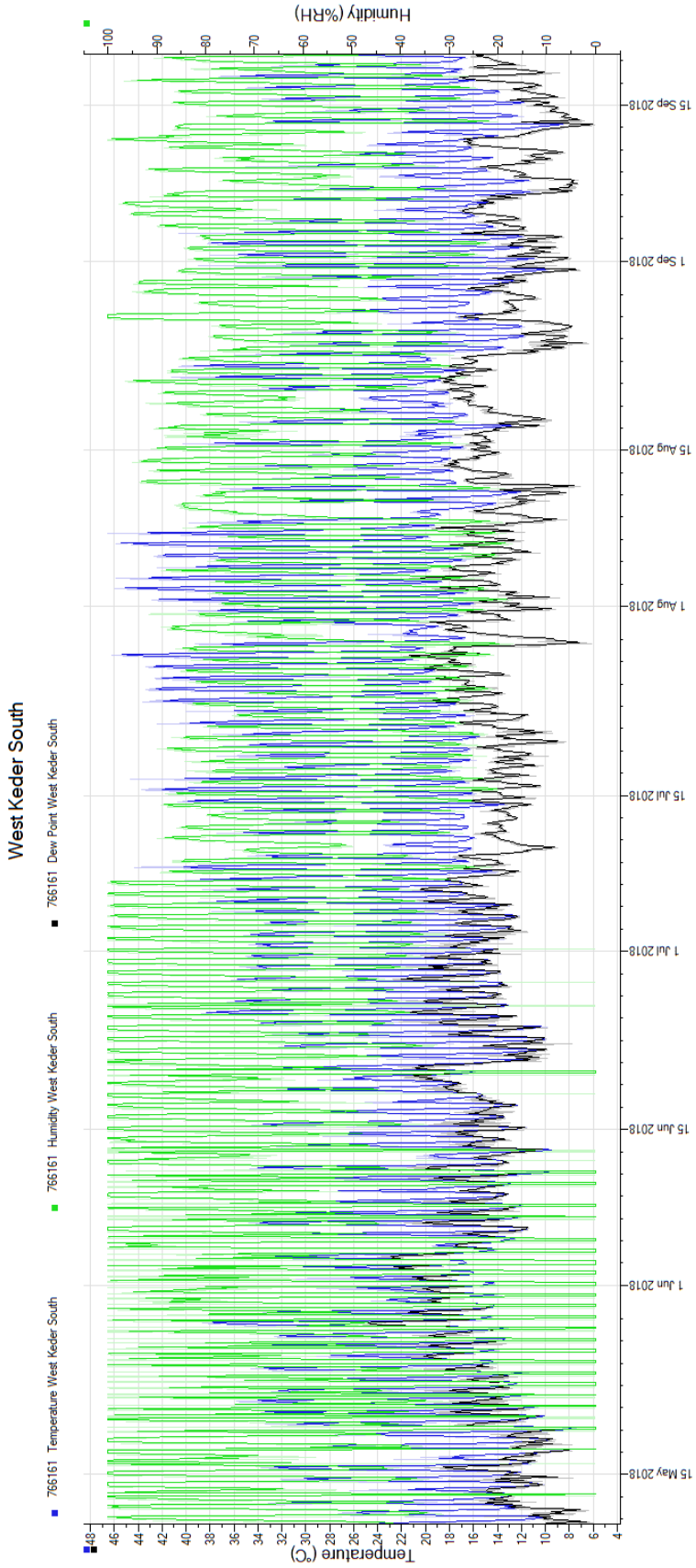
GT110251	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT110257	Broccoli	*Brassica oleracea var. italica*	DH
GT110266	Kohl Rabi	*Brassica oleracea var. gongylodes*	DH
GT110275	Kale	*Brassica oleracea var. acephala*	DH
GT120144	Broccoli	*Brassica oleracea var. italica*	S4
GT120152	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120155	Kale	*Brassica oleracea var. acephala*	DH
GT120160	Calabrese	*Brassica oleracea var. italica*	DH
GT120162	Calabrese	*Brassica oleracea var. italica*	DH
GT120163	Calabrese	*Brassica oleracea var. italica*	DH
GT120164	Broccoli	*Brassica oleracea var. italica*	DH
GT120168	Brussels Sprouts	*Brassica oleracea var. gemmifera*	DH
GT120170	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120179	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120182	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120192	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120194	Calabrese	*Brassica oleracea var. italica*	DH
GT120195	Calabrese	*Brassica oleracea var. italica*	DH
GT120198	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120201	Calabrese	*Brassica oleracea var. italica*	DH
GT120205	Kohl Rabi	*Brassica oleracea var. gongylodes*	S4
GT120208	Broccoli	*Brassica oleracea var. italica*	DH
GT120210	Cabbage	*Brassica oleracea var. capitata f. alba*	DH
GT120211	Calabrese	*Brassica oleracea var. italica*	DH
GT120213	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120214	Broccoli	*Brassica oleracea var. italica*	DH
GT120215	Broccoli	*Brassica oleracea var. italica*	DH
GT120218	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120222	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120223	Cauliflower	*Brassica oleracea var. botrytis*	DH
GT120225	Calabrese	*Brassica oleracea var. italica*	DH
GT120226	Cabbage	*Brassica oleracea var. capitata f. sabauda*	DH
GT120231	Calabrese	*Brassica oleracea var. italica*	DH
GT120233	Broccoli	*Brassica oleracea var. italica*	S4
GT120234	Broccoli	*Brassica oleracea var. italica*	S4
GT120235	Broccoli	*Brassica oleracea var. italica*	DH

S.2A: List of conditions and corresponding traits investigated using associative transcriptomics.

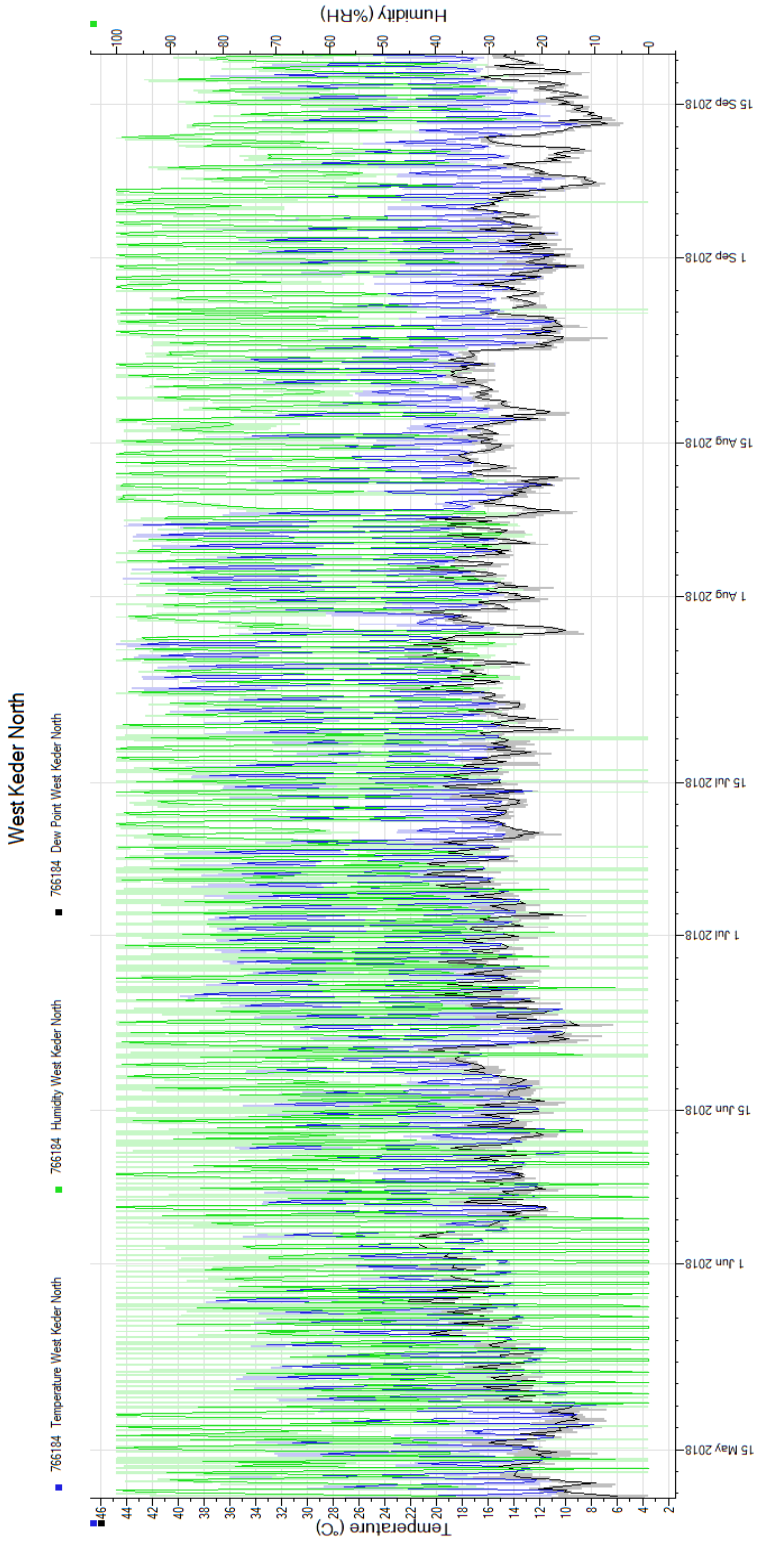
Condition	Trait
10 wk pre-growth, 12 wk vernalisation, 5°C	Days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 5°C	Days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 5°C	Difference between days to BBCH51 and days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 10°C	Days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 10°C	Days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 10°C	Difference between days to BBCH51 and days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 15°C	Days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 15°C	Days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 15°C	Difference between days to BBCH51 and days to BBCH60
6 wk pre-growth, 12 wk vernalisation, 5°C	Days to BBCH51
6 wk pre-growth, 12 wk vernalisation, 5°C	Days to BBCH60
6 wk pre-growth, 12 wk vernalisation, 5°C	Difference between days to BBCH51 and days to BBCH60
6 wk pre-growth, 12 wk vernalisation, 10°C	Days to BBCH51
6 wk pre-growth, 12 wk vernalisation, 10°C	Days to first BBCH60
6 wk pre-growth, 12 wk vernalisation, 10°C	Difference between days to BBCH51 and days to BBCH60
10 wk pre-growth, 6 wk vernalisation, 5°C	Days to BBCH51
10 wk pre-growth, 6 wk vernalisation, 5°C	Days to BBCH60
10 wk pre-growth, 6 wk vernalisation, 5°C	Difference between days to BBCH51 and days to BBCH60
10 wk pre-growth, 6 wk vernalisation, 10°C	Days to BBCH51
10 wk pre-growth, 6 wk vernalisation, 10°C	Days to BBCH60
10 wk pre-growth, 6 wk vernalisation, 10°C	Difference between days to BBCH51 and days to BBCH60
10 wk pre-growth, 6 wk vernalisation, 15°C	Days to BBCH51
10 wk pre-growth, 6 wk vernalisation, 15°C	Days to BBCH60
10 wk pre-growth, 6 wk vernalisation, 15°C	Difference between days to BBCH51 and days to BBCH60
No vernalisation	Days to BBCH51
No vernalisation	Days to first BBCH60
No vernalisation	Difference between days to BBCH51 and days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 5°C, 6 wk pre-growth, 12 wk vernalisation, 5°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 5°C, 6 wk pre-growth, 12 wk vernalisation, 5°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 10°C, 6 wk pre-growth, 12 wk vernalisation, 10°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 10°C, 6 wk pre-growth, 12 wk vernalisation, 10°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 5°C, 10 wk pre-growth, 6 wk vernalisation, 5°C	Difference in days to BBCH51

10 wk pre-growth, 12 wk vernalisation, 5°C, 10 wk pre-growth, 6 wk vernalisation, 5°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 10°C, 10 wk pre-growth, 6 wk vernalisation, 10°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 10°C, 10 wk pre-growth, 6 wk vernalisation, 10°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 15°C, 10 wk pre-growth, 6 wk vernalisation, 15°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 15°C, 10 wk pre-growth, 6 wk vernalisation, 15°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 5°C, 10 wk pre-growth, 12 wk vernalisation, 10°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 5°C, 10 wk pre-growth, 12 wk vernalisation, 10°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 5°C, 10 wk pre-growth, 12 wk vernalisation, 15°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 5°C, 10 wk pre-growth, 12 wk vernalisation, 15°C	Difference in days to BBCH60
10 wk pre-growth, 12 wk vernalisation, 10°C, 10 wk pre-growth, 12 wk vernalisation, 15°C	Difference in days to BBCH51
10 wk pre-growth, 12 wk vernalisation, 10°C, 10 wk pre-growth, 12 wk vernalisation, 15°C	Difference in days to BBCH60

S.3A: South Polytunnel temperature and humidity data recorded using TinyTag. The polytunnel was split into two halves for the purpose of recording environmental data.



S.4A: North Polytunnel temperature and humidity data recorded using TinyTag. The polytunnel was split into two halves for the purpose of recording environmental data.



S.5A: Phenotyping results, mean days to BBCH51 (DTB) and days to BBCH60 (DTF) under all treatments tested.

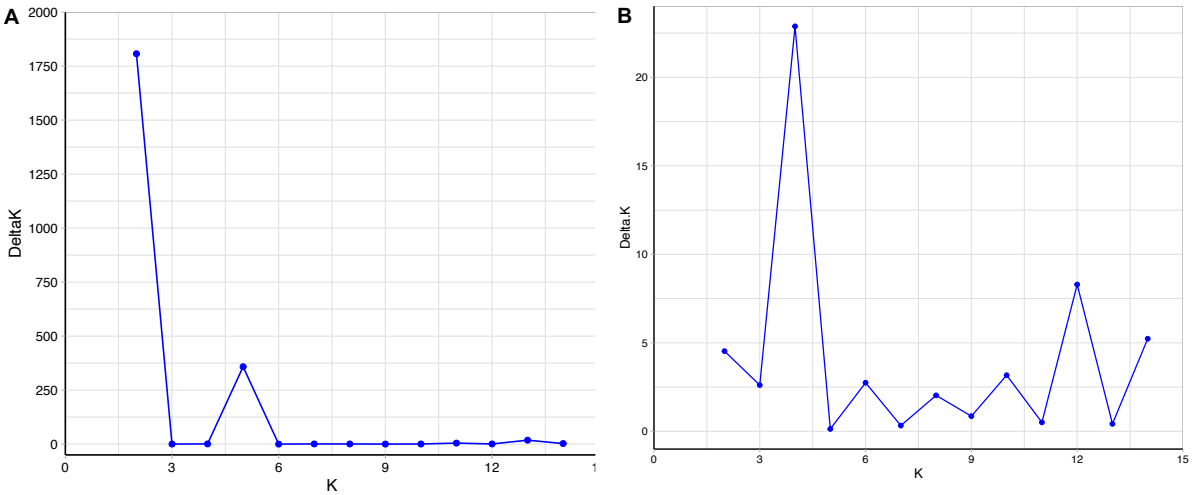
Condition	10 wk pre-growth, 12 wk vernalisation						6 wk pre-growth, 12 wk vernalisation				10 wk pre-growth, 6 wk vernalisation						Non-vernalised	
Accession	5DTB	5DTF	10DTB	10DTF	15DTB	15DTF	5DTB	5DTF	10DTB	10DTF	5DTB	5DTF	10DTB	10DTF	15DTB	15DTF	DTB	DTF
DJ3290	12.33	30.67	7.00	24.00	1.00	27.00	11.33	126.00	13.67	37.67	13.33	31.67	8.67	41.67	42.50	126.00	176.00	196.00
GK040099	9.00	37.33	2.33	30.67	17.00	40.00	126.00	126.00	6.00	126.00	22.33	44.33	30.33	42.33	21.50	48.50	170.00	196.00
GT050381	-118.67	-104.67	-118.67	-105.00	-119.33	-105.00	-113.00	-70.33	-113.00	-70.33	-94.00	-85.33	-94.00	-85.00	-93.00	-85.00	31.33	37.67
GT060867	-105.00	-99.00	126.00	126.00	-104.50	-99.00	-86.00	-79.00	-86.00	-79.50	-72.00	-57.33	-70.67	-59.67	-69.67	-57.67	37.00	53.33
GT060871	-3.33	11.67	-2.33	12.67	38.67	54.33	2.00	126.00	8.67	21.50	10.00	24.00	15.67	27.33	37.33	54.00	76.33	86.67
GT061882	7.00	34.50	1.00	25.50	2.33	31.50	21.50	75.67	126.00	126.00	48.67	126.00	-7.50	43.00	29.33	119.50	132.00	189.00
GT070804	14.00	25.00	39.67	57.33	54.33	64.00	22.33	37.00	35.50	52.00	23.00	38.33	20.00	31.50	76.33	90.00	165.00	190.00
GT080341	1.33	15.50	-35.33	6.00	-12.33	12.67	1.00	26.33	-23.00	9.00	-49.67	-29.33	-46.00	26.00	-48.00	-18.50	64.33	88.67
GT080486	5.67	16.67	1.33	13.00	3.00	126.00	4.33	19.67	9.00	19.33	6.00	22.67	11.67	26.00	126.00	126.00	68.33	80.67
GT080693	0.33	16.67	-11.00	24.67	-35.00	21.67	4.00	23.00	-24.33	6.50	5.00	17.33	18.00	31.00	-9.33	29.67	78.67	104.67
GT080713	6.33	24.33	-21.33	8.33	0.33	126.00	6.33	24.00	-18.33	126.00	15.33	28.00	-5.00	28.33	-0.33	30.67	102.67	148.00
GT080760	12.00	126.00	5.00	16.67	8.00	24.50	12.33	126.00	9.00	126.00	15.50	126.00	12.33	28.67	18.67	32.67	182.67	187.00
GT080767	-119.33	-93.67	-116.67	-94.00	-119.00	-92.33	-29.33	3.33	-73.00	-43.50	126.00	126.00	-66.33	64.67	-67.67	-54.50	48.33	54.67
GT080843	5.00	25.00	-24.67	13.00	-9.33	2.00	11.00	27.67	-21.33	14.50	14.33	31.00	2.67	21.67	4.00	24.00	85.33	102.67
GT080849	5.33	126.00	-24.00	9.67	-17.33	-2.00	6.67	26.00	-3.67	20.00	14.00	29.50	-2.33	22.33	-1.00	23.00	87.67	108.33
GT080869	6.00	21.67	-20.00	12.67	12.67	29.67	-2.67	24.67	-0.33	25.67	-1.67	56.00	-5.33	29.00	0.67	42.67	108.00	146.00
GT080876	6.33	27.33	-25.33	117.50	117.50	126.00	126.00	126.00	126.00	126.00	63.00	126.00	36.00	114.67	51.50	123.33	127.67	192.67
GT080891	2.33	9.67	-21.67	1.00	-23.00	38.33	-0.50	19.50	-43.33	-25.00	-22.50	0.67	-51.00	13.67	126.00	66.00	61.67	76.33
GT081062	126.00	126.00	126.00	126.00	126.00	126.00	1.67	50.00	-38.33	25.50	-47.00	126.00	-49.33	126.00	-46.67	126.00	111.00	146.67
GT081103	3.67	19.00	-14.33	4.67	-23.67	1.00	-1.67	25.67	-9.00	12.50	5.33	18.33	-4.33	19.00	-13.00	19.00	83.33	95.00
GT081140	6.33	126.00	-18.33	16.50	-24.67	11.67	1.33	28.67	-4.00	18.50	7.00	29.00	-8.33	24.00	2.00	41.00	80.00	101.33
GT081150	-1.67	23.00	-14.00	15.00	-0.50	54.33	6.67	27.00	126.00	126.00	11.00	35.67	-0.33	36.67	25.67	40.67	87.33	108.67
GT081395	9.33	29.00	12.00	38.00	24.67	49.00	9.50	44.50	-16.33	126.00	27.67	46.50	9.00	38.67	45.00	58.00	168.67	172.00
GT090058	86.50	122.67	126.00	126.00	120.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	124.00	126.00	124.33	126.00	190.33	196.00
GT090341	14.33	35.33	-0.33	26.67	11.00	32.33	14.67	126.00	-0.50	126.00	21.33	43.67	21.00	41.67	37.50	57.00	182.00	196.00
GT100062	2.00	126.00	-6.67	45.67	-39.00	-8.33	126.00	126.00	-34.33	3.33	-28.67	14.67	-45.00	-23.33	-28.00	0.00	66.00	80.33
GT100065	-47.00	12.67	-79.00	-34.00	-73.67	-52.67	-43.00	3.67	-53.67	-13.67	-59.33	14.33	-59.00	-45.00	-58.33	-43.00	58.00	111.33
GT100067	-12.00	34.50	-55.67	83.00	-65.67	-56.67	126.00	126.00	126.00	126.00	-16.33	34.67	-44.50	-25.00	126.00	126.00	60.67	74.33
GT100120	4.33	19.67	-24.67	8.33	-18.67	-0.67	2.33	26.67	-27.67	5.67	-13.50	33.67	-41.50	126.00	-33.67	126.00	69.33	94.67
GT110206	10.00	83.00	-4.00	117.33	45.33	114.00	38.00	66.50	47.50	120.67	42.33	93.67	46.33	77.67	27.67	126.00	108.33	175.00
GT110221	-118.67	-112.00	-119.00	-112.33	-118.67	-112.00	-92.67	-83.67	-93.00	-84.33	-84.67	-75.33	-84.33	-75.00	-84.00	-74.33	30.00	39.67
GT110244	0.67	32.00	5.00	19.33	-27.00	18.67	6.00	126.00	-16.00	39.00	13.33	33.00	-17.33	38.00	18.00	43.33	88.00	107.33
GT110251	24.50	126.00	62.33	126.00	126.00	126.00	31.00	50.33	124.00	126.00	12.67	37.67	-5.33	33.00	122.50	126.00	145.33	196.00
GT110257	14.00	32.67	0.33	25.00	6.33	27.00	25.33	53.00	25.00	126.00	16.00	33.33	3.33	27.67	28.33	49.33	87.50	101.00
GT110266	19.33	28.67	19.33	26.67	31.50	48.00	19.67	32.00	20.67	32.67	29.67	44.33	34.00	47.00	126.00	126.00	196.00	196.00
GT110275	17.67	29.67	26.00	39.67	30.00	50.50	16.67	39.67	24.33	49.50	26.33	47.67	32.33	53.67	126.00	126.00	183.67	196.00
GT120144	126.00	126.00	-23.00	3.50	6.67	28.50	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	110.33	123.67
GT120155	13.67	28.67	18.67	68.00	126.00	126.00	22.00	126.00	26.00	42.00	57.00	66.50	126.00	126.00	126.00	126.00	196.00	196.00
GT120160	17.00	35.00	24.67	37.00	126.00	126.00	126.00	126.00	126.00	126.00	31.67	48.33	126.00	126.00	126.00	126.00	196.00	196.00
GT120162	8.00	25.67	-1.50	12.00	-27.67	1.67	13.00	126.00	2.33	20.33	16.00	29.67	5.67	25.33	12.00	26.67	80.67	99.67
GT120163	6.33	23.67	-19.00	6.33	-21.33	15.33	6.67	23.67	10.00	18.33	14.33	31.67	-10.33	28.33	-27.33	61.00	82.33	100.33
GT120164	22.33	34.00	-15.67	14.00	-18.67	40.33	126.00	126.00	13.50	126.00	-11.50	43.33	20.67	32.33	21.67	34.33	97.33	108.00
GT120168	16.00	25.67	22.00	39.67	126.00	126.00	22.33	33.67	25.33	44.33	25.33	41.50	126.00	126.00	126.00	126.00	196.00	196.00
GT120170	39.67	53.00	44.50	60.50	126.00	126.00	18.33	45.33	19.67	47.67	33.00	50.00	34.00	63.33	126.00	121.00	184.33	196.00
GT120179	5.33	24.67	-27.33	10.00	-15.33	19.67	6.50	26.50	-13.33	14.33	14.33	32.33	2.33	29.67	-6.67	39.33	144.33	146.33
GT120182	8.00	23.00	13.33	32.00	26.50	126.00	5.67	26.67	-1.67	126.00	12.33	35.33	18.00	126.00	48.67	119.67	169.67	195.67
GT120192	2.00	30.67	-37.33	3.00	-4.00	126.00	7.50	126.00	-27.33	13.00	11.00	35.00	-23.00	28.00	11.67	32.67	110.00	139.00
GT120194	4.33	12.67	-32.67	0.67	-40.33	-18.67	-11.50	17.00	-21.67	-1.67	-27.00	2.33	-39.67	14.50	-22.00	10.50	62.33	196.00
GT120195	0.67	22.50	-35.00	15.33	-65.33	126.00	8.33	126.00	-25.00	22.00	0.33	23.00	-24.67	19.67	-13.67	8.67	66.00	89.00
GT120198	-94.00	126.00	-93.67	126.00	-93.33	126.00	-11.00	126.00	-57.00	5.00	-59.00	-40.67	-58.67	-41.33	-57.67	-40.67	49.67	196.00
GT120201	0.00	29.33	0.33	15.67	-4.67	22.33	8.67	27.67	-27.33	26.00	12.00	31.00	-12.67	29.00	23.33	36.00	84.33	112.33
GT120208	37.67	48.67	32.33	41.50	105.50	120.00	44.00	55.00	44.33	59.33	12.00	51.67	40.67	59.00	68.50	126.00	196.00	196.00
GT120210	5.00	14.00	-16.00	0.33	-23.67	-4.67	6.67	22.00	-7.33	3.00	14.33	27.67	4.67	21.50	7.33	22.67	111.67	123.33
GT120211	126.00	126.00	-45.00	13.33	-45.33	-6.33	9.00	27.00	-26.67	5.50	-53.33	126.00	-42.67	-4.67	9.33	126.00	65.00	83.00
GT120213	8.00	24.50	-24.67	7.33	-9.00	18.33	126.00	126.00	-27.67	126.00	14.67	29.00	-24.67	19.33	-5.00	24.67	115.00	127.67
GT120214	126.00	126.00	25.50	126.00	40.50	88.33	126.00	126.00	-12.67	126.00	75.00	126.00	77.50	117.67	-44.50	-19.00	152.67	185.00
GT120215	4.67	21.67	3.33	17.00	16.33	32.67	13.33	43.00	3.50	23.00	11.00	23.67	4.33	24.00	20.00	35.00	82.33	95.00
GT120218	1.33	32.33	-24.00	22.00	-24.33	33.00	8.00	126.00	-23.33	126.00	16.33	31.50	-24.67	30.33	11.67	43.33	103.67	138.67
GT120222	-89.67	126.00	-87.00	126.00	-85.67	30.50	9.50	126.00	-16.33	126.00	-49.00	126.00	-53.00	126.00	-50.33	126.00	64.67	196.00
GT120223	3.00	19.33	-19.67	126.00	-12.00	45.33	10.00	43.50	4.50	126.00	13.67	22.33	-0.67	34.67	17.33	46.00	98.33	122.33
GT120225	37.00	43.50	-44.33	39.67	-26.33	42.50	35.67	70.00	66.67	104.67	28.33	76.00	4.50	126.00	-34.00	-15.67	102.67	155.67
GT120226	18.00	32.33	17.67	38.67	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	126.00	166.33	171.33
GT120231	-2.00	11.50	1.00	15.00	7.67	24.67	11.33	30.33	-4.33	16.00	13.33	24.33	7.33	30.00	-46.00	37.00	106.00	147.67
GT120235	-0.33	15.33	-39.00	-3.00	17.67	126.00	-0.67	23.50	-25.00	35.50	-23.00	-1.33	-34.33	-19.00	-42.33	-30.67	63.33	80.67

S.6A: Analysis of the more stringently selected SNP data set with the Bayesian clustering algorithms implemented in the program STRUCTURE, identified four population clusters.

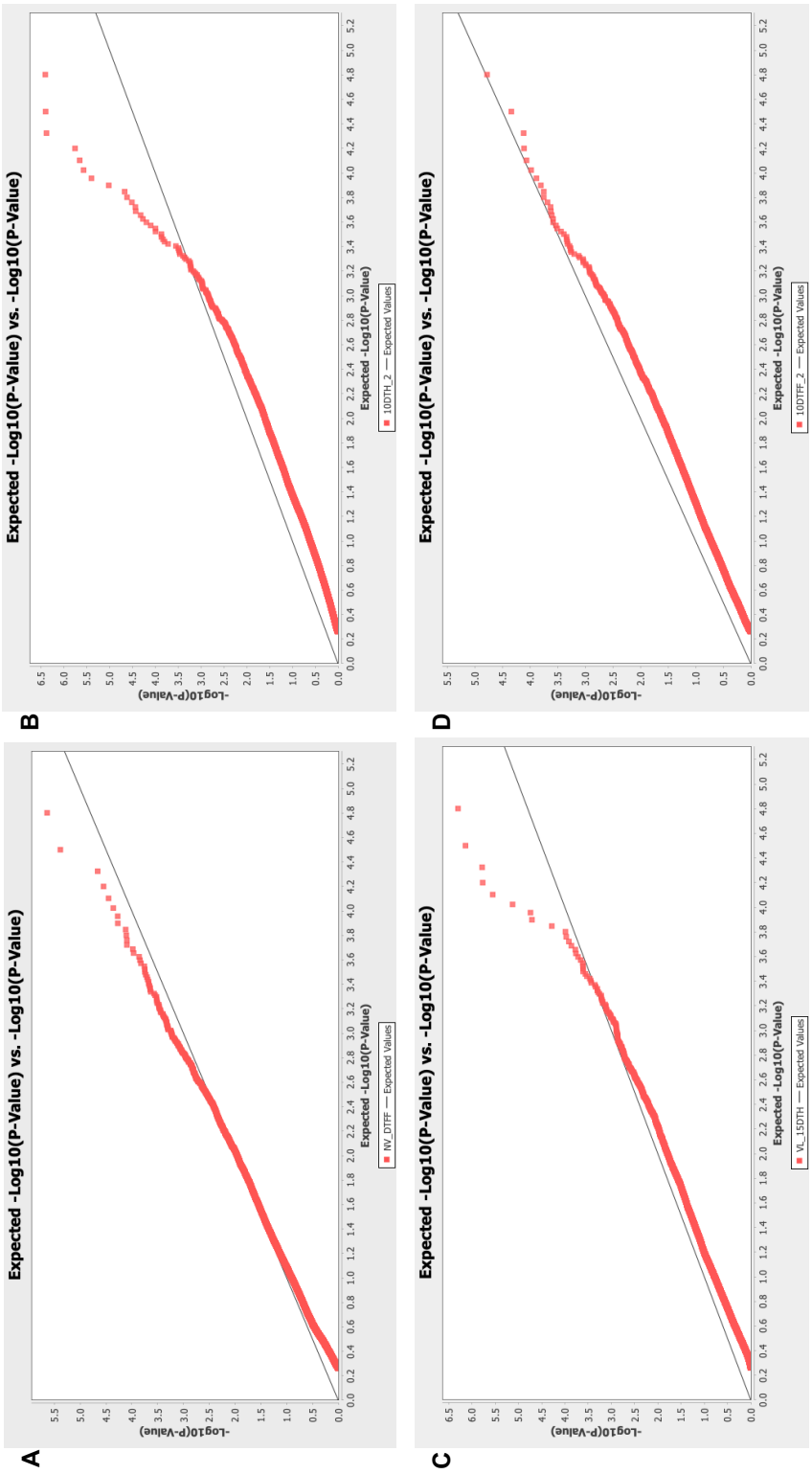
Genotype	CropType	Q1	Q2	Q3	Q4
DJ3290	Brussels Sprouts	0.002	0.002	0.002	0.995
GK040099	Cauliflower	0.002	0.006	0.606	0.385
GT050381	Kale	0.995	0.001	0.001	0.003
GT060867	Kale	0.934	0.002	0.002	0.062
GT060871	Broccoli	0.015	0.003	0.002	0.979
GT061882	Cauliflower	0.01	0.005	0.001	0.984
GT070804	Cabbage	0.035	0.002	0.012	0.952
GT080341	Broccoli	0.01	0.005	0.006	0.979
GT080486	Kale	0.022	0.002	0.005	0.97
GT080693	Calabrese	0.005	0.988	0.004	0.003
GT080713	Cauliflower	0.001	0.002	0.996	0.001
GT080760	Kale	0.008	0.002	0.002	0.988
GT080767	Kale	0.99	0.002	0.002	0.006
GT080843	Calabrese	0.001	0.997	0.001	0.001
GT080849	Broccoli	0.001	0.997	0.001	0.001
GT080869	Cauliflower	0.019	0.008	0.374	0.598
GT080876	Cauliflower	0.003	0.014	0.976	0.006
GT080891	Kale	0.018	0.045	0.19	0.747
GT081012	Kale	0.181	0.022	0.017	0.78
GT081062	Cauliflower	0.004	0.003	0.907	0.087
GT081103	Broccoli	0.003	0.005	0.003	0.989
GT081140	Broccoli	0.002	0.981	0.011	0.005
GT081150	Broccoli	0.001	0.997	0.001	0.001
GT081395	Broccoli	0.021	0.043	0.019	0.916
GT090058	Cabbage	0.003	0.002	0.001	0.994
GT090341	Cauliflower	0.001	0.001	0.997	0.001
GT100062	Broccoli	0.002	0.968	0.02	0.009
GT100065	Broccoli	0.004	0.355	0.223	0.419
GT100067	Kale	0.992	0.001	0.003	0.003
GT100120	Broccoli	0.001	0.899	0.098	0.002
GT110206	Cauliflower	0.001	0.002	0.992	0.004
GT110221	Kale	0.992	0.002	0.004	0.002
GT110244	Broccoli	0.009	0.843	0.142	0.006
GT110251	Cauliflower	0.001	0.005	0.987	0.006
GT110257	Calabrese	0.001	0.975	0.001	0.023
GT110266	KohlRabi	0.002	0.003	0.992	0.003
GT110275	Kale	0.002	0.009	0.002	0.986
GT120144	Broccoli	0.013	0.139	0.161	0.687
GT120152	Cauliflower	0.002	0.003	0.565	0.43
GT120155	Kale	0.007	0.003	0.009	0.98
GT120160	Calabrese	0.004	0.974	0.019	0.003
GT120162	Calabrese	0.003	0.979	0.015	0.002
GT120163	Calabrese	0.004	0.126	0.003	0.868
GT120164	Broccoli	0.3	0.028	0.005	0.668
GT120168	Brussels Sprouts	0.001	0.001	0.998	0.001
GT120170	Cauliflower	0.018	0.011	0.882	0.089
GT120179	Cauliflower	0.003	0.005	0.459	0.533
GT120182	Cauliflower	0.002	0.088	0.33	0.58
GT120192	Cauliflower	0.001	0.001	0.997	0.001
GT120194	Calabrese	0.002	0.845	0.15	0.004
GT120195	Calabrese	0.001	0.995	0.002	0.002
GT120198	Cauliflower	0.02	0.004	0.473	0.504
GT120201	Calabrese	0.002	0.991	0.004	0.002

GT120205	KohlRabi	0.008	0.065	0.018	0.909
GT120208	Broccoli	0.004	0.184	0.26	0.553
GT120210	Cabbage	0.007	0.009	0.049	0.935
GT120211	Calabrese	0.001	0.996	0.002	0.001
GT120213	Cauliflower	0.001	0.001	0.997	0.001
GT120214	Broccoli	0.004	0.649	0.012	0.335
GT120215	Broccoli	0.002	0.629	0.005	0.363
GT120218	Cauliflower	0.002	0.002	0.993	0.003
GT120222	Cauliflower	0.004	0.696	0.213	0.087
GT120223	Cauliflower	0.004	0.137	0.336	0.523
GT120225	Calabrese	0.001	0.981	0.016	0.002
GT120226	Cabbage	0.003	0.004	0.008	0.985
GT120231	Calabrese	0.001	0.997	0.001	0.002
GT120233	Broccoli	0.001	0.018	0.002	0.978
GT120234	Broccoli	0.003	0.006	0.002	0.989
GT120235	Broccoli	0.001	0.003	0.007	0.989

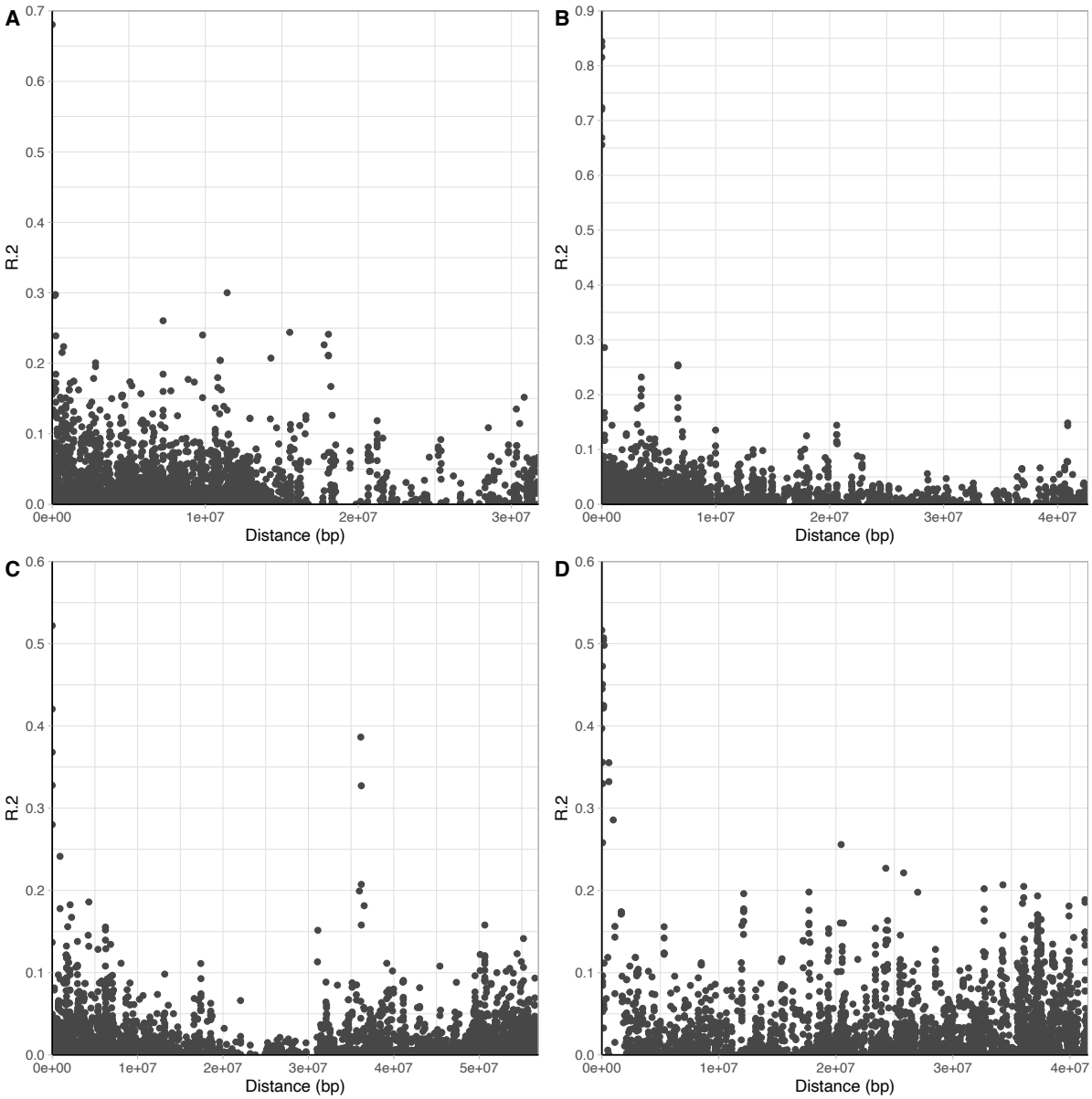
S.7A: *ΔK* based on rate of change of LnP, Maxima indicates the *ΔK* that best explains the population structure. Plots produced using STRUCTURE Harvester output. A) *ΔK* values calculated for SNPs with MAF > 0.05, *K* = 5. B) *ΔK* values for biallelic SNPs, MAF > 0.05, one SNP per gene, > 500 bp apart, *K* = 4.



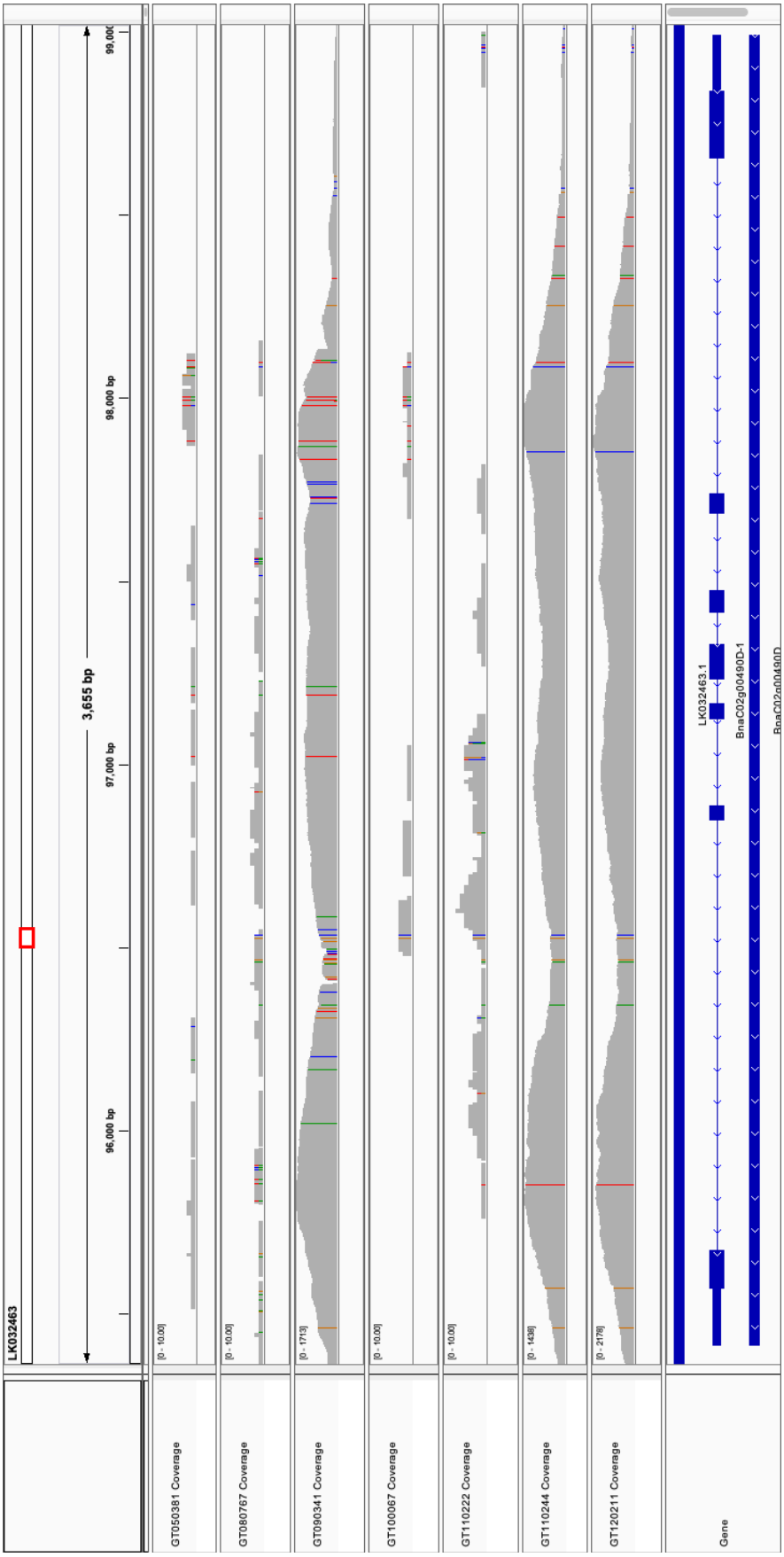
S.8A: Quantile-Quantile Plots for SNP associations with A) days to BBCH60 under NV conditions. GLM, with Q matrix correction for population structure B) days to BBCH51 after a six-week pre-growth and 10 ºC vernalisation for twelve-weeks. GLM, with Q matrix correction for population structure C) The difference in days to BBCH51 following exposure to 5 ºC and 15 ºC vernalisation for six-weeks, after exposure to a ten-week pre-growth. GLM with Q matrix correction for population structure D) Days to BBCH60 after exposure to a six-week pre-growth and twelve weeks vernalisation at 10 ºC. GLM with PCA correction for population structure.



S.9A: Linkage disequilibrium decay. A) Bo8g089990.1:453:T, miR172D candidate. B) Bo9g179000.1:2589:G, *ELF6* candidate. C) Bo7g026810.1:124:G, *FIP1* candidate. D) Bo7g104810.1:204:T, miR172D candidate.



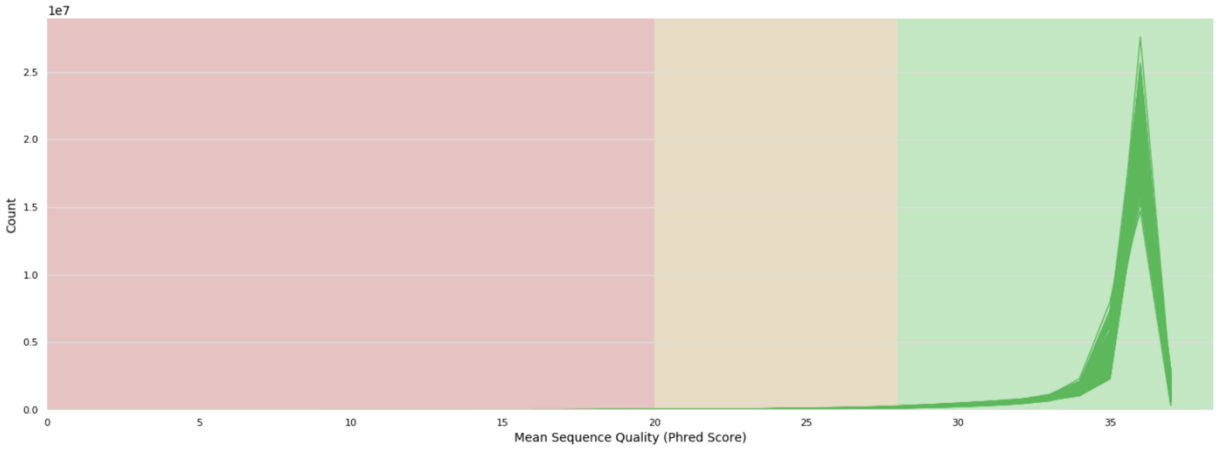
S.10A: Mapping *BoFLC.C02* using Darmor-*bzh* as a reference. Four rapid cycling accessions and three representative accessions for the rest of the population.



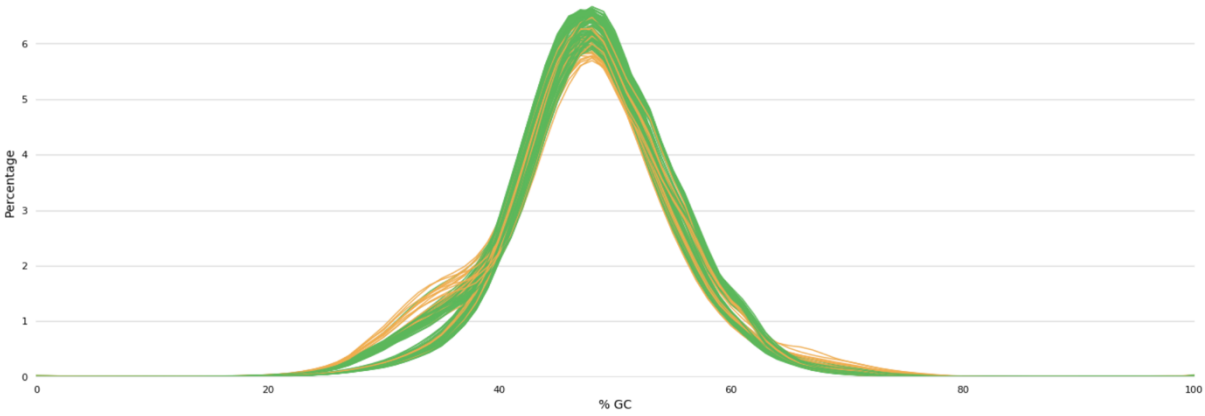
Appendix B

Supplemental Figures and Tables for Chapter 3.

S.1B: Per sequence quality scores for raw reads from DH1012 time series datas, calculated as phred scores. Plot exported from FastQC.



S.2B: The average percentage GC content of raw reads from DH1012 time series data. Plot exported from FastQC.



.

S.3B: Statistics and sample information for alignment to the *B. oleracea* pantranscriptome, including total number of mapped reads and the percentage of uniquely mapping reads.

Sample	Tissue	Days post sowing	Total Reads	Total Passed QC	Mapped	Secondary Alignments	Uniquely mapping reads (%)	Duplicates
D12A1_1	Apex	14	28.79850	28.79850	28.79850	0.32634	98.87	0
D12A1_2	Apex	14	41.82442	41.82442	41.82442	0.50142	98.80	0
D12A1_3	Apex	14	38.19637	38.19637	38.19637	0.46230	98.79	0
D12A2_1	Apex	21	37.12621	37.12621	37.12621	0.41047	98.89	0
D12A2_2	Apex	21	34.35695	34.35695	34.35695	0.39200	98.86	0
D12A2_3	Apex	21	28.32630	28.32630	28.32630	0.31097	98.90	0
D12A3_1	Apex	23	35.34086	35.34086	35.34086	0.36673	98.96	0
D12A3_2	Apex	23	28.20646	28.20646	28.20646	0.34120	98.79	0
D12A3_3	Apex	23	32.57111	32.57111	32.57111	0.37316	98.85	0
D12A4_1	Apex	25	39.67119	39.67119	39.67119	0.42314	98.93	0
D12A4_2	Apex	25	32.19996	32.19996	32.19996	0.39248	98.78	0
D12A4_3	Apex	25	34.52749	34.52749	34.52749	0.37915	98.90	0
D12A5_1	Apex	27	28.03620	28.03620	28.03620	0.31028	98.89	0
D12A5_2	Apex	27	36.38789	36.38789	36.38789	0.37549	98.97	0
D12A5_3	Apex	27	31.88438	31.88438	31.88438	0.32061	98.99	0
D12A6_1	Apex	29	30.28378	30.28378	30.28378	0.33182	98.90	0
D12A6_2	Apex	29	41.52115	41.52115	41.52115	0.40355	99.03	0
D12A6_3	Apex	29	38.88190	38.88190	38.88190	0.41591	98.93	0
D12A7_1	Apex	31	35.03331	35.03331	35.03331	0.37259	98.94	0
D12A7_2	Apex	31	41.78699	41.78699	41.78699	0.43546	98.96	0
D12A7_3	Apex	31	36.41705	36.41705	36.41705	0.37309	98.98	0
D12A8_1	Apex	33	28.20609	28.20609	28.20609	0.31568	98.88	0
D12A8_2	Apex	33	38.80240	38.80240	38.80240	0.38334	99.01	0
D12A8_3	Apex	33	28.77043	28.77043	28.77043	0.29811	98.96	0
D12A9_1	Apex	35	34.88133	34.88133	34.88133	0.37152	98.93	0
D12A9_2	Apex	35	33.53968	33.53968	33.53968	0.36986	98.90	0
D12A9_3	Apex	35	28.84914	28.84914	28.84914	0.30749	98.93	0
D12A10_1	Apex	37	31.78579	31.78579	31.78579	0.32401	98.98	0
D12A10_2	Apex	37	33.13036	33.13036	33.13036	0.33380	98.99	0
D12A10_3	Apex	37	39.10147	39.10147	39.10147	0.37657	99.04	0
D12A11_1	Apex	39	34.34203	34.34203	34.34203	0.33731	99.02	0
D12A11_2	Apex	39	32.96629	32.96629	32.96629	0.37721	98.86	0
D12A11_3	Apex	39	35.22374	35.22374	35.22374	0.32807	99.07	0
D12A12_1	Apex	41	38.22753	38.22753	38.22753	0.38749	98.99	0

D12A12_2	Apex	41	30.81297	30.81297	30.81297	0.33744	98.90	0
D12A12_3	Apex	41	30.84554	30.84554	30.84554	0.35441	98.8	0
D12A13_1	Apex	43	33.94929	33.94929	33.94929	0.35333	98.96	0
D12A13_2	Apex	43	31.78631	31.78631	31.78631	0.33788	98.94	0
D12A13_3	Apex	43	28.39988	28.39988	28.39988	0.28362	99.00	0
D12A14_1	Apex	49	31.22079	31.22079	31.22079	0.33386	98.93	0
D12A14_2	Apex	49	29.73611	29.73611	29.73611	0.32239	98.92	0
D12A14_3	Apex	49	29.99443	29.99443	29.99443	0.33474	98.88	0
D12A15_1	Apex	51	32.36678	32.36678	32.36678	0.33500	98.9	0
D12A15_2	Apex	51	30.26610	30.26610	30.26610	0.34223	98.87	0
D12A15_3	Apex	51	30.72752	30.72752	30.72752	0.35420	98.85	0
D12L1_1	Leaf	14	32.80569	32.80569	32.80569	0.50240	98.47	0
D12L1_2	Leaf	14	30.33382	30.33382	30.33382	0.42681	98.59	0
D12L1_3	Leaf	14	33.83192	33.83192	33.83192	0.46120	98.64	0
D12L2_1	Leaf	21	32.52644	32.52644	32.52644	0.55920	98.28	0
D12L2_2	Leaf	21	35.30969	35.30969	35.30969	0.51654	98.54	0
D12L2_3	Leaf	21	32.41207	32.41207	32.41207	0.52923	98.37	0
D12L3_1	Leaf	23	30.38839	30.38839	30.38839	0.33082	98.91	0
D12L3_2	Leaf	23	32.53887	32.53887	32.53887	0.52502	98.39	0
D12L3_3	Leaf	23	35.94523	35.94523	35.94523	0.42410	98.82	0
D12L4_1	Leaf	25	35.34263	35.34263	35.34263	0.39771	98.87	0
D12L4_2	Leaf	25	38.80441	38.80441	38.80441	0.38853	99.00	0
D12L4_3	Leaf	25	39.17334	39.17334	39.17334	0.48295	98.77	0
D12L5_1	Leaf	27	34.96221	34.96221	34.96221	0.33170	99.05	0
D12L5_2	Leaf	27	35.92711	35.92711	35.92711	0.46115	98.72	0
D12L5_3	Leaf	27	42.01868	42.01868	42.01868	0.51654	98.77	0
D12L6_1	Leaf	29	33.56864	33.56864	33.56864	0.31924	99.05	0
D12L6_2	Leaf	29	38.62531	38.62531	38.62531	0.46617	98.79	0
D12L6_3	Leaf	29	35.29602	35.29602	35.29602	0.38356	98.91	0
D12L7_1	Leaf	31	37.28358	37.28358	37.28358	0.37841	98.99	0
D12L7_2	Leaf	31	36.18712	36.18712	36.18712	0.41583	98.85	0
D12L7_3	Leaf	31	38.86291	38.86291	38.86291	0.41757	98.93	0
D12L8_1	Leaf	33	34.88736	34.88736	34.88736	0.37499	98.93	0
D12L8_2	Leaf	33	44.76589	44.76589	44.76589	0.46957	98.95	0
D12L8_3	Leaf	33	35.49624	35.49624	35.49624	0.37041	98.96	0
D12L9_1	Leaf	35	34.85191	34.85191	34.85191	0.32997	99.05	0
D12L9_2	Leaf	35	32.90721	32.90721	32.90721	0.37230	98.87	0
D12L9_3	Leaf	35	32.92973	32.92973	32.92973	0.43295	98.69	0
D12SL_1	Seedling	7	33.17722	33.17722	33.17722	0.51844	98.44	0
D12SL_2	Seedling	7	32.41696	32.41696	32.41696	0.64883	98.00	0
D12SL_3	Seedling	7	41.99071	41.99071	41.99071	0.56799	98.65	0

S.4B: Statistics and sample information for alignment to the *B. oleracea* pangenome, including total number of mapped reads and the percentage of uniquely mapping reads.

Sample Name	Tissue	Days post sowing	Total Reads	Total Passed QC	Mapped	Secondary Alignments	Uniquely mapping reads (%)	Duplicates
D12A1_1	Apex	14	49.48876	49.48876	49.48876	4.13597	91.64	0
D12A1_2	Apex	14	69.37275	69.37275	69.37275	5.15627	92.57	0
D12A1_3	Apex	14	62.22636	62.22636	62.22636	4.55012	92.69	0
D12A2_1	Apex	21	65.1356	65.1356	65.1356	4.82654	92.59	0
D12A2_2	Apex	21	58.24296	58.24296	58.24296	4.24862	92.71	0
D12A2_3	Apex	21	49.63325	49.63325	49.63325	5.26938	89.38	0
D12A3_1	Apex	23	60.90856	60.90856	60.90856	4.5665	92.50	0
D12A3_2	Apex	23	48.59516	48.59516	48.59516	4.60158	90.53	0
D12A3_3	Apex	23	54.87447	54.87447	54.87447	4.04901	92.62	0
D12A4_1	Apex	25	67.35588	67.35588	67.35588	4.94431	92.66	0
D12A4_2	Apex	25	56.26419	56.26419	56.26419	4.45799	92.08	0
D12A4_3	Apex	25	58.47655	58.47655	58.47655	4.10377	92.98	0
D12A5_1	Apex	27	48.48373	48.48373	48.48373	4.11478	91.51	0
D12A5_2	Apex	27	60.61548	60.61548	60.61548	4.35114	92.82	0
D12A5_3	Apex	27	55.54328	55.54328	55.54328	4.12651	92.57	0
D12A6_1	Apex	29	50.39316	50.39316	50.39316	3.7468	92.56	0
D12A6_2	Apex	29	70.19563	70.19563	70.19563	5.02171	92.85	0
D12A6_3	Apex	29	68.27211	68.27211	68.27211	6.30088	90.77	0
D12A7_1	Apex	31	59.79942	59.79942	59.79942	4.59148	92.32	0
D12A7_2	Apex	31	69.39821	69.39821	69.39821	5.10463	92.64	0
D12A7_3	Apex	31	60.15867	60.15867	60.15867	4.51464	92.50	0
D12A8_1	Apex	33	47.41854	47.41854	47.41854	3.60204	92.40	0
D12A8_2	Apex	33	67.26589	67.26589	67.26589	5.07001	92.46	0
D12A8_3	Apex	33	50.32644	50.32644	50.32644	3.47976	93.09	0
D12A9_1	Apex	35	56.64134	56.64134	56.64134	4.34771	92.32	0
D12A9_2	Apex	35	56.98444	56.98444	56.98444	4.44688	92.20	0
D12A9_3	Apex	35	49.4351	49.4351	49.4351	3.47274	92.98	0
D12A10_1	Apex	37	55.12742	55.12742	55.12742	4.16218	92.45	0
D12A10_2	Apex	37	55.29245	55.29245	55.29245	3.94243	92.87	0
D12A10_3	Apex	37	65.76222	65.76222	65.76222	4.70648	92.84	0
D12A11_1	Apex	39	57.4931	57.4931	57.4931	4.25197	92.60	0
D12A11_2	Apex	39	55.22749	55.22749	55.22749	4.03225	92.70	0
D12A11_3	Apex	39	59.27694	59.27694	59.27694	4.13274	93.03	0
D12A12_1	Apex	41	64.28137	64.28137	64.28137	4.77085	92.58	0
D12A12_2	Apex	41	51.30098	51.30098	51.30098	3.92144	92.36	0
D12A12_3	Apex	41	53.17114	53.17114	53.17114	4.42053	91.69	0

D12A13_1	Apex	43	57.05192	57.05192	57.05192	4.24526	92.56	0
D12A13_2	Apex	43	52.6021	52.6021	52.6021	3.93106	92.53	0
D12A13_3	Apex	43	48.82938	48.82938	48.82938	3.71232	92.40	0
D12A14_1	Apex	49	57.92363	57.92363	57.92363	4.29803	92.58	0
D12A14_2	Apex	49	53.45319	53.45319	53.45319	3.91637	92.67	0
D12A14_3	Apex	49	53.94745	53.94745	53.94745	4.09855	92.40	0
D12A15_1	Apex	51	59.57311	59.57311	59.57311	4.3649	92.67	0
D12A15_2	Apex	51	53.46897	53.46897	53.46897	3.89592	92.71	0
D12A15_3	Apex	51	52.06286	52.06286	52.06286	3.74235	92.81	0
D12L1_1	Leaf	14	58.82783	58.82783	58.82783	4.96622	91.56	0
D12L1_2	Leaf	14	48.61364	48.61364	48.61364	3.9174	91.94	0
D12L1_3	Leaf	14	59.06602	59.06602	59.06602	4.79588	91.88	0
D12L2_1	Leaf	21	50.88165	50.88165	50.88165	5.28767	89.61	0
D12L2_2	Leaf	21	57.81842	57.81842	57.81842	6.07218	89.50	0
D12L2_3	Leaf	21	54.02704	54.02704	54.02704	5.47122	89.87	0
D12L3_1	Leaf	23	51.16253	51.16253	51.16253	5.37106	89.50	0
D12L3_2	Leaf	23	53.96035	53.96035	53.96035	5.50046	89.81	0
D12L3_3	Leaf	23	62.85233	62.85233	62.85233	7.50564	88.06	0
D12L4_1	Leaf	25	60.17179	60.17179	60.17179	6.72864	88.82	0
D12L4_2	Leaf	25	63.49581	63.49581	63.49581	6.36924	89.97	0
D12L4_3	Leaf	25	65.82424	65.82424	65.82424	6.80662	89.66	0
D12L5_1	Leaf	27	57.17172	57.17172	57.17172	6.17748	89.19	0
D12L5_2	Leaf	27	58.59475	58.59475	58.59475	6.26798	89.30	0
D12L5_3	Leaf	27	69.77714	69.77714	69.77714	7.51373	89.23	0
D12L6_1	Leaf	29	55.26404	55.26404	55.26404	5.93345	89.26	0
D12L6_2	Leaf	29	61.75877	61.75877	61.75877	6.51966	89.44	0
D12L6_3	Leaf	29	59.07697	59.07697	59.07697	6.48202	89.03	0
D12L7_1	Leaf	31	63.99063	63.99063	63.99063	6.20422	90.30	0
D12L7_2	Leaf	31	60.7528	60.7528	60.7528	5.50611	90.94	0
D12L7_3	Leaf	31	65.33671	65.33671	65.33671	6.40834	90.19	0
D12L8_1	Leaf	33	59.21612	59.21612	59.21612	4.86687	91.78	0
D12L8_2	Leaf	33	76.21435	76.21435	76.21435	6.63517	91.29	0
D12L8_3	Leaf	33	59.52939	59.52939	59.52939	6.55698	88.99	0
D12L9_1	Leaf	35	59.57775	59.57775	59.57775	5.07563	91.48	0
D12L9_2	Leaf	35	54.67639	54.67639	54.67639	4.26826	92.19	0
D12L9_3	Leaf	35	54.78805	54.78805	54.78805	5.14334	90.61	0
D12SL_1	Seedling	7	58.99591	58.99591	58.99591	6.77514	88.52	0
D12SL_2	Seedling	7	65.01688	65.01688	65.01688	10.72647	83.50	0
D12SL_3	Seedling	7	70.1355	70.1355	70.1355	7.46926	89.35	0

S.5B: Enriched gene ontology terms for the upregulated DEGs from the 14 d 35 d apex comparison. Terms with a direct involvement in the floral transition highlighted in green.

Identifier	Type	Log2-Enrichment Fold	P-Value	Description
GO:0048367	BP	0.93	1.47E-09	shoot system development
GO:0010468	BP	0.66	2.15E-09	regulation of gene expression
GO:0006355	BP	0.71	2.48E-09	regulation of transcription, DNA-templated
GO:1903506	BP	0.71	2.48E-09	regulation of nucleic acid-templated transcription
GO:2001141	BP	0.71	2.56E-09	regulation of RNA biosynthetic process
GO:0051252	BP	0.7	3.70E-09	regulation of RNA metabolic process
GO:0003677	MF	0.68	5.54E-09	DNA binding
GO:0031326	BP	0.66	6.00E-09	regulation of cellular biosynthetic process
GO:0065007	BP	0.43	6.90E-09	biological regulation
GO:2000112	BP	0.67	9.63E-09	regulation of cellular macromolecule biosynthetic process
GO:0009889	BP	0.65	9.73E-09	regulation of biosynthetic process
GO:0090567	BP	1.11	9.92E-09	reproductive shoot system development
GO:0010556	BP	0.67	1.36E-08	regulation of macromolecule biosynthetic process
GO:0019219	BP	0.67	1.44E-08	regulation of nucleobase-containing compound metabolic process
GO:0009908	BP	1.12	1.59E-08	flower development
GO:0006351	BP	0.66	2.73E-08	transcription, DNA-templated
GO:0097659	BP	0.66	2.73E-08	nucleic acid-templated transcription
GO:0032774	BP	0.66	2.97E-08	RNA biosynthetic process
GO:0060255	BP	0.6	3.23E-08	regulation of macromolecule metabolic process
GO:0003700	MF	1	4.54E-08	transcription factor activity, sequence-specific DNA binding
GO:0001071	MF	0.99	5.70E-08	nucleic acid binding transcription factor activity
GO:0051171	BP	0.61	1.20E-07	regulation of nitrogen compound metabolic process
GO:0050789	BP	0.42	2.33E-07	regulation of biological process
GO:0050794	BP	0.45	3.02E-07	regulation of cellular process
GO:0034654	BP	0.6	3.84E-07	nucleobase-containing compound biosynthetic process
GO:0019222	BP	0.53	7.63E-07	regulation of metabolic process
GO:0031323	BP	0.55	1.56E-06	regulation of cellular metabolic process
GO:0080090	BP	0.56	1.67E-06	regulation of primary metabolic process
GO:0043565	MF	0.95	2.06E-06	sequence-specific DNA binding
GO:1901700	BP	0.62	2.28E-06	response to oxygen-containing compound
GO:0019438	BP	0.53	7.61E-06	aromatic compound biosynthetic process
GO:0000977	MF	2.47	1.16E-05	RNA polymerase II regulatory region sequence-specific DNA binding
GO:0001012	MF	2.47	1.16E-05	RNA polymerase II regulatory region DNA binding
GO:0018130	BP	0.52	1.60E-05	heterocycle biosynthetic process
GO:1901362	BP	0.49	3.94E-05	organic cyclic compound biosynthetic process

GO:0000976	MF	1.75	5.82E-05	transcription regulatory region sequence-specific DNA binding
GO:0016070	BP	0.45	2.60E-04	RNA metabolic process
GO:1990837	MF	1.55	3.20E-04	sequence-specific double-stranded DNA binding
GO:0043531	MF	1.51	3.52E-04	ADP binding
GO:0031328	BP	1.12	3.55E-04	positive regulation of cellular biosynthetic process
GO:0050896	BP	0.3	3.95E-04	response to stimulus
GO:0097159	MF	0.25	4.08E-04	organic cyclic compound binding
GO:0048522	BP	0.84	4.80E-04	positive regulation of cellular process
GO:0048518	BP	0.72	5.46E-04	positive regulation of biological process
GO:0010033	BP	0.47	6.17E-04	response to organic substance
GO:1901363	MF	0.25	6.20E-04	heterocyclic compound binding
GO:0051173	BP	1.04	6.64E-04	positive regulation of nitrogen compound metabolic process
GO:0001101	BP	0.58	6.78E-04	response to acid chemical
GO:0048831	BP	1.19	6.85E-04	regulation of shoot system development
GO:1902680	BP	1.19	7.26E-04	positive regulation of RNA biosynthetic process
GO:1903508	BP	1.19	7.26E-04	positive regulation of nucleic acid-templated transcription
GO:0046983	MF	0.81	8.61E-04	protein dimerization activity
GO:0042221	BP	0.39	1.12E-03	response to chemical
GO:0009891	BP	1.06	1.32E-03	positive regulation of biosynthetic process
GO:0051254	BP	1.14	1.67E-03	positive regulation of RNA metabolic process
GO:0048523	BP	0.78	1.88E-03	negative regulation of cellular process
GO:0031325	BP	0.97	1.94E-03	positive regulation of cellular metabolic process
GO:0006950	BP	0.39	1.97E-03	response to stress
GO:0006952	BP	0.66	2.63E-03	defence response
GO:0007154	BP	0.54	2.65E-03	cell communication
GO:0048731	BP	0.46	2.76E-03	system development
GO:0045944	BP	2.03	2.96E-03	positive regulation of transcription from RNA polymerase II promoter
GO:0003676	MF	0.34	3.21E-03	nucleic acid binding
GO:0010557	BP	1.07	3.29E-03	positive regulation of macromolecule biosynthetic process
GO:0090304	BP	0.38	3.48E-03	nucleic acid metabolic process
GO:0045893	BP	1.15	3.50E-03	positive regulation of transcription, DNA-templated
GO:0044212	MF	1.04	5.77E-03	transcription regulatory region DNA binding
GO:0048519	BP	0.62	6.69E-03	negative regulation of biological process
GO:0007165	BP	0.56	6.74E-03	signal transduction
GO:0005488	MF	0.13	7.13E-03	binding
GO:0009862	BP	2.6	7.56E-03	systemic acquired resistance, salicylic acid mediated signalling pathway
GO:0044700	BP	0.55	8.77E-03	single organism signalling
GO:0023052	BP	0.55	9.36E-03	signalling
GO:0000975	MF	1.01	9.60E-03	regulatory region DNA binding
GO:0001067	MF	1.01	9.60E-03	regulatory region nucleic acid binding

GO:0071229	BP	0.8	1.01E-02	cellular response to acid chemical
GO:0045935	BP	1.02	1.03E-02	positive regulation of nucleobase-containing compound metabolic process
GO:0051240	BP	1.44	1.03E-02	positive regulation of multicellular organismal process
GO:0043621	MF	1.63	1.08E-02	protein self-association
GO:0051239	BP	0.76	1.30E-02	regulation of multicellular organismal process
GO:2000022	BP	2.9	1.37E-02	regulation of jasmonic acid mediated signalling pathway
GO:0010628	BP	1.04	1.56E-02	positive regulation of gene expression
GO:2000026	BP	0.78	1.58E-02	regulation of multicellular organismal development
GO:0009739	BP	1.22	1.80E-02	response to gibberellin
GO:0051094	BP	1.35	2.03E-02	positive regulation of developmental process
GO:0005634	CC	0.26	2.12E-02	nucleus
GO:0043170	BP	0.22	2.37E-02	macromolecule metabolic process
GO:0009893	BP	0.83	2.60E-02	positive regulation of metabolic process
GO:1901701	BP	0.67	3.11E-02	cellular response to oxygen-containing compound
GO:0010243	BP	1.28	3.18E-02	response to organonitrogen compound
GO:0003690	MF	1.08	3.32E-02	double-stranded DNA binding
GO:0051716	BP	0.4	3.49E-02	cellular response to stimulus
GO:0010604	BP	0.87	4.75E-02	positive regulation of macromolecule metabolic process

S.6B: Enriched gene ontology terms for the upregulated DEGs from the 14 d 51 d apex comparison. Terms with a direct involvement in the floral transition highlighted in green.

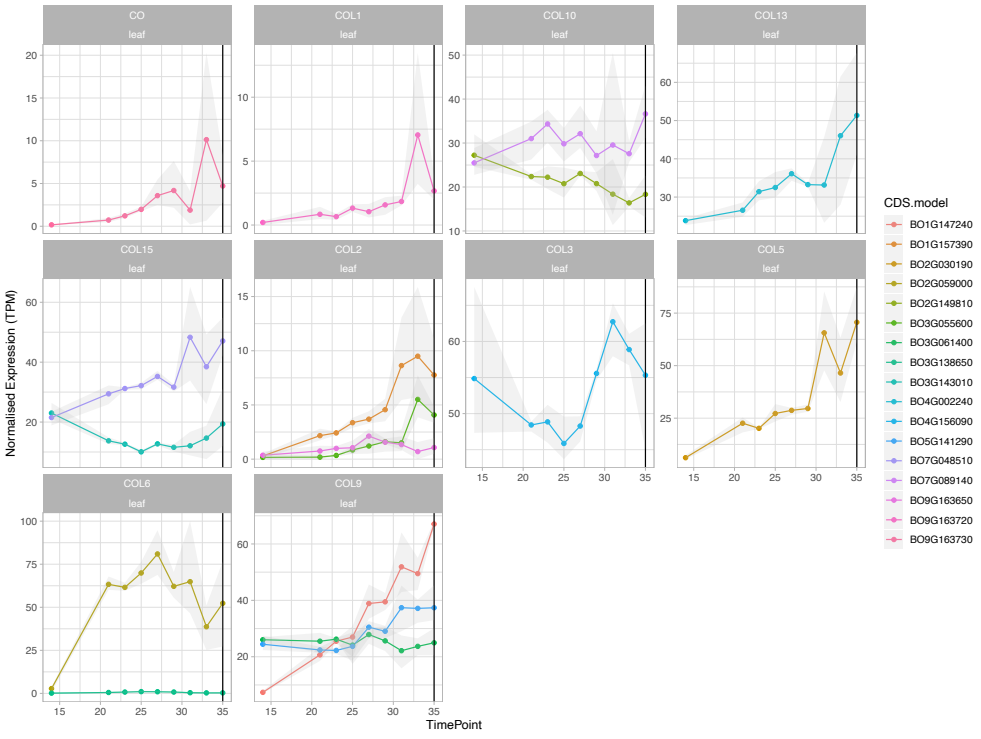
Identifier	Type	Log2-Enrichment Fold	P-Value	Description
GO:0000977	MF	2.46	3.86E-18	RNA polymerase II regulatory region sequence-specific DNA binding
GO:0001012	MF	2.46	3.86E-18	RNA polymerase II regulatory region DNA binding
GO:0003700	MF	0.86	1.76E-15	transcription factor activity, sequence-specific DNA binding
GO:0001071	MF	0.86	2.90E-15	nucleic acid binding transcription factor activity
GO:0031323	BP	0.47	1.00E-12	regulation of cellular metabolic process
GO:0000976	MF	1.65	1.34E-12	transcription regulatory region sequence-specific DNA binding
GO:0006355	BP	0.53	1.72E-12	regulation of transcription, DNA-templated
GO:1903506	BP	0.53	1.72E-12	regulation of nucleic acid-templated transcription
GO:2001141	BP	0.53	1.82E-12	regulation of RNA biosynthetic process
GO:0051252	BP	0.52	2.99E-12	regulation of RNA metabolic process
GO:0019219	BP	0.5	9.03E-12	regulation of nucleobase-containing compound metabolic process
GO:0003677	MF	0.49	2.33E-11	DNA binding
GO:0009889	BP	0.47	4.36E-11	regulation of biosynthetic process
GO:0080090	BP	0.45	4.93E-11	regulation of primary metabolic process
GO:0031326	BP	0.46	1.36E-10	regulation of cellular biosynthetic process
GO:0051171	BP	0.45	1.45E-10	regulation of nitrogen compound metabolic process
GO:0010556	BP	0.46	4.51E-10	regulation of macromolecule biosynthetic process
GO:0006351	BP	0.46	5.01E-10	transcription, DNA-templated
GO:0097659	BP	0.46	5.01E-10	nucleic acid-templated transcription
GO:0010033	BP	0.43	5.50E-10	response to organic substance
GO:0032774	BP	0.46	5.80E-10	RNA biosynthetic process
GO:2000112	BP	0.46	5.80E-10	regulation of cellular macromolecule biosynthetic process
GO:0019222	BP	0.4	7.93E-10	regulation of metabolic process
GO:0045944	BP	1.91	1.47E-09	positive regulation of transcription from RNA polymerase II promoter
GO:0090567	BP	0.77	2.43E-09	reproductive shoot system development
GO:1990837	MF	1.38	2.96E-09	sequence-specific double-stranded DNA binding
GO:0009908	BP	0.77	4.16E-09	flower development
GO:0016602	CC	3.19	5.10E-09	CCAAT-binding factor complex
GO:0009891	BP	0.95	6.97E-09	positive regulation of biosynthetic process
GO:1901700	BP	0.45	1.01E-08	response to oxygen-containing compound
GO:0042221	BP	0.35	1.32E-08	response to chemical
GO:0009725	BP	0.45	1.54E-08	response to hormone

GO:0031328	BP	0.95	1.97E-08	positive regulation of cellular biosynthetic process
GO:0009719	BP	0.45	2.22E-08	response to endogenous stimulus
GO:0019438	BP	0.39	4.60E-08	aromatic compound biosynthetic process
GO:0034654	BP	0.41	5.50E-08	nucleobase-containing compound biosynthetic process
GO:0080167	BP	1.09	6.35E-08	response to karrikin
GO:1901362	BP	0.38	6.36E-08	organic cyclic compound biosynthetic process
GO:0043565	MF	0.69	6.55E-08	sequence-specific DNA binding
GO:0001101	BP	0.49	7.85E-08	response to acid chemical
GO:0060255	BP	0.38	8.22E-08	regulation of macromolecule metabolic process
GO:0044212	MF	0.97	1.36E-07	transcription regulatory region DNA binding
GO:0010557	BP	0.95	1.39E-07	positive regulation of macromolecule biosynthetic process
GO:0000975	MF	0.95	2.14E-07	regulatory region DNA binding
GO:0001067	MF	0.95	2.14E-07	regulatory region nucleic acid binding
GO:0050794	BP	0.29	2.58E-07	regulation of cellular process
GO:0018130	BP	0.37	2.71E-07	heterocycle biosynthetic process
GO:0010468	BP	0.39	3.23E-07	regulation of gene expression
GO:0065007	BP	0.24	2.69E-06	biological regulation
GO:1902680	BP	0.93	4.27E-06	positive regulation of RNA biosynthetic process
GO:1903508	BP	0.93	4.27E-06	positive regulation of nucleic acid-templated transcription
GO:0045935	BP	0.87	5.94E-06	positive regulation of nucleobase-containing compound metabolic process
GO:0050896	BP	0.22	6.65E-06	response to stimulus
GO:0044699	BP	0.17	7.83E-06	single-organism process
GO:0048522	BP	0.63	9.29E-06	positive regulation of cellular process
GO:0048367	BP	0.51	1.41E-05	shoot system development
GO:0051254	BP	0.88	2.04E-05	positive regulation of RNA metabolic process
GO:0046983	MF	0.61	2.20E-05	protein dimerization activity
GO:0033993	BP	0.51	2.36E-05	response to lipid
GO:0014070	BP	0.72	3.67E-05	response to organic cyclic compound
GO:1905421	BP	1.74	3.87E-05	regulation of plant organ morphogenesis
GO:0010143	BP	2.31	3.87E-05	cutin biosynthetic process
GO:0050789	BP	0.23	5.04E-05	regulation of biological process
GO:0009751	BP	0.93	9.65E-05	response to salicylic acid
GO:1900618	BP	1.81	9.67E-05	regulation of shoot system morphogenesis
GO:0031325	BP	0.71	1.08E-04	positive regulation of cellular metabolic process
GO:0006357	BP	1.32	1.25E-04	regulation of transcription from RNA polymerase II promoter
GO:0048518	BP	0.5	1.29E-04	positive regulation of biological process
GO:0048437	BP	0.83	1.32E-04	floral organ development
GO:0051173	BP	0.73	1.44E-04	positive regulation of nitrogen compound metabolic process
GO:0009416	BP	0.52	1.61E-04	response to light stimulus

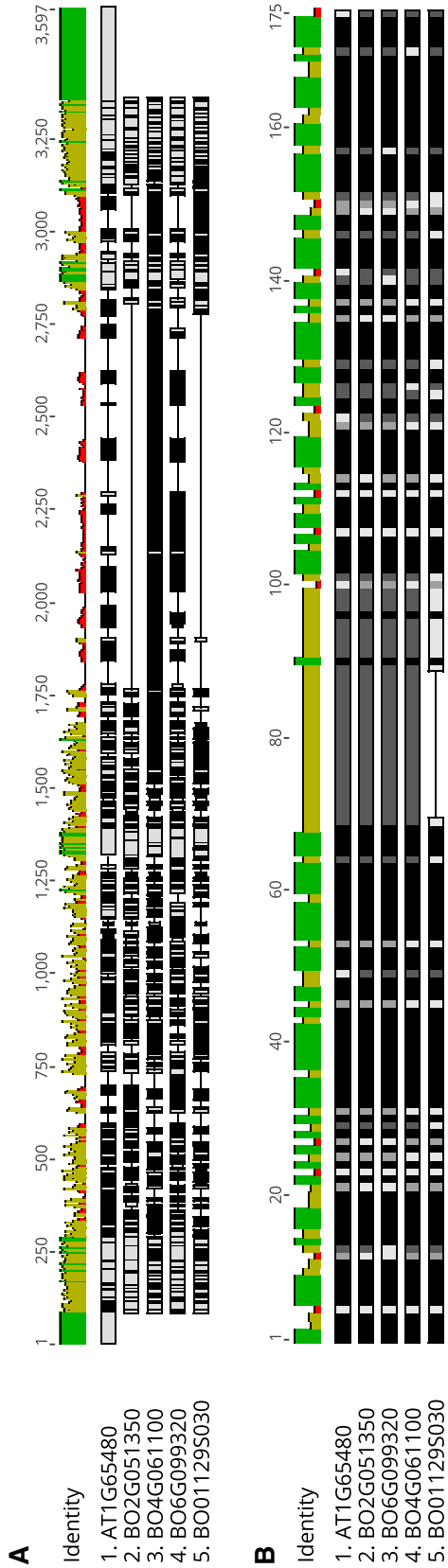
GO:1901568	BP	1.74	2.49E-04	fatty acid derivative metabolic process
GO:0009893	BP	0.64	5.25E-04	positive regulation of metabolic process
GO:0009628	BP	0.29	5.77E-04	response to abiotic stimulus
GO:0048831	BP	0.81	5.97E-04	regulation of shoot system development
GO:0042335	BP	1.59	6.64E-04	cuticle development
GO:0010604	BP	0.69	7.48E-04	positive regulation of macromolecule metabolic process
GO:0009314	BP	0.49	7.73E-04	response to radiation
GO:0003690	MF	0.85	1.15E-03	double-stranded DNA binding
GO:0010166	BP	1.72	1.16E-03	wax metabolic process
GO:0045893	BP	0.8	1.57E-03	positive regulation of transcription, DNA-templated
GO:0045962	BP	2.44	2.20E-03	positive regulation of development, heterochronic
GO:0090575	CC	1.9	2.35E-03	RNA polymerase II transcription factor complex
GO:0010025	BP	1.7	2.91E-03	wax biosynthetic process
GO:0048523	BP	0.5	3.14E-03	negative regulation of cellular process
GO:0006366	BP	1.09	3.36E-03	transcription from RNA polymerase II promoter
GO:0048827	BP	0.53	3.52E-03	phyllome development
GO:1901570	BP	1.67	3.86E-03	fatty acid derivative biosynthetic process
GO:0009607	BP	0.42	3.93E-03	response to biotic stimulus
GO:0022603	BP	1	4.91E-03	regulation of anatomical structure morphogenesis
GO:0048646	BP	0.77	5.76E-03	anatomical structure formation involved in morphogenesis
GO:0043207	BP	0.41	6.04E-03	response to external biotic stimulus
GO:0051707	BP	0.41	6.04E-03	response to other organism
GO:2000024	BP	1.19	7.47E-03	regulation of leaf development
GO:0009415	BP	0.56	8.82E-03	response to water
GO:0009733	BP	0.58	9.50E-03	response to auxin
GO:0019748	BP	0.71	1.03E-02	secondary metabolic process
GO:1905392	BP	0.55	1.12E-02	plant organ morphogenesis
GO:0004668	MF	3.31	1.24E-02	protein-arginine deiminase activity
GO:0009753	BP	0.76	1.26E-02	response to jasmonic acid
GO:0016491	MF	0.35	1.29E-02	oxidoreductase activity
GO:0009414	BP	0.55	1.34E-02	response to water deprivation
GO:1905393	BP	0.94	1.55E-02	plant organ formation
GO:0090698	BP	0.78	1.74E-02	post-embryonic plant morphogenesis
GO:0090696	BP	0.76	1.78E-02	post-embryonic plant organ development
GO:0010628	BP	0.69	1.83E-02	positive regulation of gene expression
GO:0080154	BP	3.08	2.06E-02	regulation of fertilization
GO:2000306	BP	2.18	2.14E-02	positive regulation of photomorphogenesis
GO:0008544	BP	1.62	2.30E-02	epidermis development
GO:0009913	BP	1.62	2.30E-02	epidermal cell differentiation
GO:0030855	BP	1.62	2.30E-02	epithelial cell differentiation

GO:0042659	BP	1.96	2.37E-02	regulation of cell fate specification
GO:0051704	BP	0.32	2.37E-02	multi-organism process
GO:0044798	CC	1.62	2.75E-02	nuclear transcription factor complex
GO:0046524	MF	2.89	2.84E-02	sucrose-phosphate synthase activity
GO:0032787	BP	0.58	3.14E-02	monocarboxylic acid metabolic process
GO:0006952	BP	0.38	3.53E-02	defence response
GO:0072330	BP	0.78	4.35E-02	monocarboxylic acid biosynthetic process
GO:0044710	BP	0.2	4.55E-02	single-organism metabolic process
GO:0051240	BP	0.93	4.84E-02	positive regulation of multicellular organismal process

S.7B: Expression of *CO*-like orthologues identified in *B. oleracea* DH1012 over the floral transition in the leaf tissue. For most *CO*-like orthologues, expression peaks in the days before the apex transitioned to an inflorescence. Expression measured as TPM, black line indicates day the apex had morphologically transitioned to an inflorescence meristem.



S.8B: Alignment of Arabidopsis *FT* and its *B. oleracea* orthologues A) genomic sequences and B) protein sequences. A large deletion is present in BO01129S030 that is not present in the other paralogues.



Appendix C

Supplementary Figures and Tables for Chapter 4.

S.1C: Expression data (TPM) used to construct gene regulatory networks to describe the interactions between genes.

		Expression (TPM)					
Replicate	Timepoint (days)	AP1 BO6G108600	AP1 BO6G095760	LFY BO2G161690	LFY BO3G109270	TFL1 BO0983S080	TFL1 BO9G181670
1	14	0.000	0.000	2.492	2.183	186.641	21.484
2	14	0.000	0.000	1.583	1.476	184.871	22.101
3	14	0.000	0.000	1.253	1.274	162.688	17.696
1	21	0.000	0.000	1.277	1.451	422.860	31.203
2	21	0.000	0.000	0.598	1.762	345.624	27.874
3	21	0.000	0.000	1.462	3.137	423.630	41.506
1	23	0.000	0.000	2.214	1.004	370.031	35.121
2	23	0.000	0.000	2.038	0.783	339.697	31.434
3	23	0.000	0.000	1.038	1.649	382.586	27.950
1	25	0.000	0.000	1.926	1.908	386.252	40.445
2	25	0.000	0.000	1.505	0.791	357.965	32.813
3	25	0.000	0.000	0.926	1.098	365.020	20.021
1	27	0.000	0.000	2.770	2.145	505.643	42.755
2	27	0.000	0.000	3.129	3.465	465.322	54.594
3	27	0.000	0.000	5.301	4.715	513.597	44.513
1	29	0.000	0.000	4.827	3.602	426.456	28.628
2	29	0.000	0.000	2.759	5.173	397.107	41.188
3	29	0.000	0.000	1.956	1.299	432.135	45.842
1	31	0.000	0.000	7.958	9.842	538.740	57.451
2	31	0.000	0.000	2.280	3.567	433.455	62.118
3	31	0.000	0.000	5.374	4.902	563.023	38.421
1	33	0.000	0.000	1.281	4.418	573.305	51.905
2	33	0.000	0.000	4.642	6.432	397.790	38.631
3	33	0.238	0.000	8.597	13.554	358.971	20.463
1	35	0.000	0.000	16.816	16.935	432.630	34.607
2	35	0.000	0.000	6.187	7.975	406.080	45.568
3	35	2.878	1.748	14.984	16.305	342.588	28.233
1	37	7.333	9.420	26.320	27.106	308.892	24.074
2	37	0.000	0.000	5.268	7.518	327.902	20.492
3	37	0.000	0.000	4.948	7.581	302.872	34.580
1	39	12.166	13.814	16.398	15.876	341.048	42.799
2	39	0.000	0.000	4.270	3.134	307.614	51.155
3	39	14.870	17.070	25.681	23.699	353.296	36.555
1	41	63.563	63.876	37.629	36.969	121.242	16.430
2	41	32.383	34.987	29.319	25.111	145.097	13.871
3	41	61.350	75.540	68.220	63.901	126.295	8.255
1	49	87.790	154.503	19.687	42.526	32.489	2.713
2	49	101.356	165.334	20.078	30.051	5.384	1.901
3	49	117.794	165.748	22.972	48.329	34.483	1.504
1	51	85.105	158.979	11.186	23.725	5.022	1.147
2	51	84.340	174.285	7.060	13.408	3.222	1.450
3	51	83.494	177.641	3.824	9.797	1.069	0.000

S.2C: Gene ontology enrichment analysis for genes with the same registration parameters (stretch, shift) in DH1012.

Identifier	Type	P-Value	Description	Stretch	Shift
GO:0043436	BP	8.06E-13	oxoacid metabolic process	1.5	-3.0
GO:0006082	BP	8.68E-13	organic acid metabolic process	1.5	-3.0
GO:0046496	BP	5.32E-12	nicotinamide nucleotide metabolic process	1.5	-3.0
GO:0006096	BP	5.32E-12	glycolytic process	1.5	-3.0
GO:0006757	BP	5.32E-12	ATP generation from ADP	1.5	-3.0
GO:0019752	BP	5.77E-12	carboxylic acid metabolic process	1.5	-3.0
GO:0009135	BP	6.10E-12	purine nucleoside diphosphate metabolic process	1.5	-3.0
GO:0009179	BP	6.10E-12	purine ribonucleoside diphosphate metabolic process	1.5	-3.0
GO:0009185	BP	6.10E-12	ribonucleoside diphosphate metabolic process	1.5	-3.0
GO:0046031	BP	6.10E-12	ADP metabolic process	1.5	-3.0
GO:0019362	BP	6.37E-12	pyridine nucleotide metabolic process	1.5	-3.0
GO:0044281	BP	1.31E-11	small molecule metabolic process	1.5	-3.0
GO:0044724	BP	1.34E-11	single-organism carbohydrate catabolic process	1.5	-3.0
GO:0072524	BP	1.40E-11	pyridine-containing compound metabolic process	1.5	-3.0
GO:0006165	BP	1.51E-11	nucleoside diphosphate phosphorylation	1.5	-3.0
GO:0009132	BP	2.18E-11	nucleoside diphosphate metabolic process	1.5	-3.0
GO:0046939	BP	2.18E-11	nucleotide phosphorylation	1.5	-3.0
GO:0006733	BP	3.43E-11	oxidoreduction coenzyme metabolic process	1.5	-3.0
GO:0006090	BP	2.14E-10	pyruvate metabolic process	1.5	-3.0
GO:0016052	BP	4.27E-10	carbohydrate catabolic process	1.5	-3.0
GO:0006732	BP	1.44E-09	coenzyme metabolic process	1.5	-3.0
GO:0051186	BP	1.45E-08	cofactor metabolic process	1.5	-3.0
GO:0044710	BP	3.36E-08	single-organism metabolic process	1.5	-3.0
GO:0046034	BP	3.71E-08	ATP metabolic process	1.5	-3.0
GO:0051775	BP	4.20E-08	response to redox state	1.5	-3.0
GO:0009144	BP	5.98E-08	purine nucleoside triphosphate metabolic process	1.5	-3.0
GO:0009205	BP	5.98E-08	purine ribonucleoside triphosphate metabolic process	1.5	-3.0
GO:0006094	BP	9.65E-08	gluconeogenesis	1.5	-3.0
GO:0072521	BP	1.27E-07	purine-containing compound metabolic process	1.5	-3.0
GO:0009199	BP	1.31E-07	ribonucleoside triphosphate metabolic process	1.5	-3.0
GO:0009126	BP	1.39E-07	purine nucleoside monophosphate metabolic process	1.5	-3.0
GO:0009167	BP	1.39E-07	purine ribonucleoside monophosphate metabolic process	1.5	-3.0

GO:0006091	BP	1.56E-07	generation of precursor metabolites and energy	1.5	-3.0
GO:0009141	BP	1.91E-07	nucleoside triphosphate metabolic process	1.5	-3.0
GO:0006006	BP	2.49E-07	glucose metabolic process	1.5	-3.0
GO:0055086	BP	2.55E-07	nucleobase-containing small molecule metabolic process	1.5	-3.0
GO:0009161	BP	3.02E-07	ribonucleoside monophosphate metabolic process	1.5	-3.0
GO:0009123	BP	3.87E-07	nucleoside monophosphate metabolic process	1.5	-3.0
GO:0032787	BP	4.46E-07	monocarboxylic acid metabolic process	1.5	-3.0
GO:0009150	BP	9.41E-07	purine ribonucleotide metabolic process	1.5	-3.0
GO:0009117	BP	1.21E-06	nucleotide metabolic process	1.5	-3.0
GO:0044699	BP	1.44E-06	single-organism process	1.5	-3.0
GO:0019319	BP	1.48E-06	hexose biosynthetic process	1.5	-3.0
GO:0006753	BP	1.52E-06	nucleoside phosphate metabolic process	1.5	-3.0
GO:0006163	BP	1.59E-06	purine nucleotide metabolic process	1.5	-3.0
GO:1901135	BP	1.65E-06	carbohydrate derivative metabolic process	1.5	-3.0
GO:0009259	BP	2.13E-06	ribonucleotide metabolic process	1.5	-3.0
GO:0044283	BP	3.24E-06	small molecule biosynthetic process	1.5	-3.0
GO:0044723	BP	5.74E-06	single-organism carbohydrate metabolic process	1.5	-3.0
GO:0010038	BP	6.65E-06	response to metal ion	1.5	-3.0
GO:0046686	BP	7.24E-06	response to cadmium ion	1.5	-3.0
GO:0019693	BP	8.20E-06	ribose phosphate metabolic process	1.5	-3.0
GO:0044763	BP	8.30E-06	single-organism cellular process	1.5	-3.0
GO:0008152	BP	8.49E-06	metabolic process	1.5	-3.0
GO:0042221	BP	1.09E-05	response to chemical	1.5	-3.0
GO:0046364	BP	1.23E-05	monosaccharide biosynthetic process	1.5	-3.0
GO:0005975	BP	1.50E-05	carbohydrate metabolic process	1.5	-3.0
GO:0044238	BP	2.78E-05	primary metabolic process	1.5	-3.0
GO:0019318	BP	3.70E-05	hexose metabolic process	1.5	-3.0
GO:0071704	BP	4.10E-05	organic substance metabolic process	1.5	-3.0
GO:0009987	BP	5.33E-05	cellular process	1.5	-3.0
GO:0010033	BP	8.48E-05	response to organic substance	1.5	-3.0
GO:0005996	BP	1.02E-04	monosaccharide metabolic process	1.5	-3.0
GO:0044712	BP	1.41E-04	single-organism catabolic process	1.5	-3.0
GO:0009744	BP	1.60E-04	response to sucrose	1.5	-3.0
GO:0042542	BP	1.68E-04	response to hydrogen peroxide	1.5	-3.0
GO:0034285	BP	2.56E-04	response to disaccharide	1.5	-3.0
GO:0044237	BP	3.80E-04	cellular metabolic process	1.5	-3.0
GO:0010035	BP	6.72E-04	response to inorganic substance	1.5	-3.0
GO:0055114	BP	1.07E-03	oxidation-reduction process	1.5	-3.0
GO:1901575	BP	1.08E-03	organic substance catabolic process	1.5	-3.0

GO:0019637	BP	1.31E-03	organophosphate metabolic process	1.5	-3.0
GO:0009056	BP	1.86E-03	catabolic process	1.5	-3.0
GO:1901564	BP	2.17E-03	organonitrogen compound metabolic process	1.5	-3.0
GO:0044711	BP	2.27E-03	single-organism biosynthetic process	1.5	-3.0
GO:0048608	BP	2.51E-03	reproductive structure development	1.5	-3.0
GO:0061458	BP	2.51E-03	reproductive system development	1.5	-3.0
GO:0006807	BP	2.83E-03	nitrogen compound metabolic process	1.5	-3.0
GO:1901360	BP	3.93E-03	organic cyclic compound metabolic process	1.5	-3.0
GO:0009651	BP	4.91E-03	response to salt stress	1.5	-3.0
GO:0009743	BP	4.92E-03	response to carbohydrate	1.5	-3.0
GO:0006725	BP	5.89E-03	cellular aromatic compound metabolic process	1.5	-3.0
GO:0006536	BP	6.76E-03	glutamate metabolic process	1.5	-3.0
GO:0046483	BP	7.15E-03	heterocycle metabolic process	1.5	-3.0
GO:0071555	BP	8.36E-03	cell wall organization	1.5	-3.0
GO:0016053	BP	8.57E-03	organic acid biosynthetic process	1.5	-3.0
GO:0046394	BP	8.57E-03	carboxylic acid biosynthetic process	1.5	-3.0
GO:0009628	BP	9.48E-03	response to abiotic stimulus	1.5	-3.0
GO:0003006	BP	1.01E-02	developmental process involved in reproduction	1.5	-3.0
GO:0009791	BP	1.04E-02	post-embryonic development	1.5	-3.0
GO:0009058	BP	1.78E-02	biosynthetic process	1.5	-3.0
GO:0006970	BP	1.92E-02	response to osmotic stress	1.5	-3.0
GO:0048316	BP	1.94E-02	seed development	1.5	-3.0
GO:0048731	BP	2.05E-02	system development	1.5	-3.0
GO:0009408	BP	2.51E-02	response to heat	1.5	-3.0
GO:0044702	BP	2.63E-02	single organism reproductive process	1.5	-3.0
GO:0045229	BP	3.01E-02	external encapsulating structure organization	1.5	-3.0
GO:0006139	BP	3.27E-02	nucleobase-containing compound metabolic process	1.5	-3.0
GO:0010154	BP	3.32E-02	fruit development	1.5	-3.0
GO:0050896	BP	3.71E-02	response to stimulus	1.5	-3.0
GO:0009225	BP	4.33E-02	nucleotide-sugar metabolic process	1.5	-3.0
GO:0034641	BP	4.79E-02	cellular nitrogen compound metabolic process	1.5	-3.0
GO:0004365	MF	2.72E-13	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity	1.5	-3.0
GO:0043891	MF	2.72E-13	glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+) (phosphorylating) activity	1.5	-3.0
GO:0008886	MF	1.30E-11	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (non-phosphorylating) activity	1.5	-3.0
GO:0051287	MF	2.35E-07	NAD binding	1.5	-3.0
GO:0003824	MF	3.65E-07	catalytic activity	1.5	-3.0

GO:0016620	MF	4.67E-06	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	1.5	-3.0
GO:0005507	MF	1.84E-05	copper ion binding	1.5	-3.0
GO:0050661	MF	2.97E-05	NADP binding	1.5	-3.0
GO:0016903	MF	6.32E-05	oxidoreductase activity, acting on the aldehyde or oxo group of donors	1.5	-3.0
GO:0048037	MF	5.35E-04	cofactor binding	1.5	-3.0
GO:0050662	MF	1.49E-03	coenzyme binding	1.5	-3.0
GO:0003989	MF	2.98E-03	acetyl-CoA carboxylase activity	1.5	-3.0
GO:0016421	MF	4.08E-03	CoA carboxylase activity	1.5	-3.0
GO:0016885	MF	4.08E-03	ligase activity, forming carbon-carbon bonds	1.5	-3.0
GO:0016491	MF	7.93E-03	oxidoreductase activity	1.5	-3.0
GO:0004489	MF	1.80E-02	methylenetetrahydrofolate reductase (NAD(P)H) activity	1.5	-3.0
GO:0004615	MF	1.80E-02	phosphomannomutase activity	1.5	-3.0
GO:0009374	MF	3.59E-02	biotin binding	1.5	-3.0
GO:0004185	MF	3.89E-02	serine-type carboxypeptidase activity	1.5	-3.0
GO:0005576	CC	1.72E-12	extracellular region	1.5	-3.0
GO:0048046	CC	4.08E-08	apoplast	1.5	-3.0
GO:0005618	CC	2.13E-06	cell wall	1.5	-3.0
GO:0030312	CC	2.13E-06	external encapsulating structure	1.5	-3.0
GO:0044464	CC	2.57E-05	cell part	1.5	-3.0
GO:0005623	CC	3.72E-05	cell	1.5	-3.0
GO:0005829	CC	1.07E-04	cytosol	1.5	-3.0
GO:0031967	CC	3.15E-04	organelle envelope	1.5	-3.0
GO:0031975	CC	3.18E-04	envelope	1.5	-3.0
GO:0009507	CC	3.97E-04	chloroplast	1.5	-3.0
GO:0009570	CC	8.05E-04	chloroplast stroma	1.5	-3.0
GO:0005737	CC	9.44E-04	cytoplasm	1.5	-3.0
GO:0009532	CC	1.08E-03	plastid stroma	1.5	-3.0
GO:0009536	CC	1.09E-03	plastid	1.5	-3.0
GO:0005622	CC	2.84E-03	intracellular	1.5	-3.0
GO:0044424	CC	3.02E-03	intracellular part	1.5	-3.0
GO:0044444	CC	6.01E-03	cytoplasmic part	1.5	-3.0
GO:0005773	CC	6.88E-03	vacuole	1.5	-3.0
GO:0043231	CC	8.20E-03	intracellular membrane-bounded organelle	1.5	-3.0
GO:0071944	CC	8.53E-03	cell periphery	1.5	-3.0
GO:0005777	CC	9.31E-03	peroxisome	1.5	-3.0
GO:0042579	CC	9.31E-03	microbody	1.5	-3.0
GO:0043227	CC	1.18E-02	membrane-bounded organelle	1.5	-3.0
GO:0005740	CC	2.10E-02	mitochondrial envelope	1.5	-3.0
GO:0043229	CC	2.99E-02	intracellular organelle	1.5	-3.0
GO:0043226	CC	3.05E-02	organelle	1.5	-3.0

GO:0044434	CC	3.56E-02	chloroplast part	1.5	-3.0
GO:0009317	CC	4.19E-02	acetyl-CoA carboxylase complex	1.5	-3.0
GO:0044435	CC	4.64E-02	plastid part	1.5	-3.0
GO:0016673	MF	3.89E-05	oxidoreductase activity, acting on a sulfur group of donors, iron-sulfur protein as acceptor	1.5	-2.2
GO:0050311	MF	3.89E-05	sulfite reductase (ferredoxin) activity	1.5	-2.2
GO:0051539	MF	4.65E-02	4 iron, 4 sulfur cluster binding	1.5	-2.2
GO:0010439	BP	9.51E-04	regulation of glucosinolate biosynthetic process	1.5	-1.4
GO:0044272	BP	1.66E-03	sulfur compound biosynthetic process	1.5	-1.4
GO:1900376	BP	6.57E-03	regulation of secondary metabolite biosynthetic process	1.5	-1.4
GO:0010675	BP	1.74E-02	regulation of cellular carbohydrate metabolic process	1.5	-1.4
GO:0042762	BP	2.26E-02	regulation of sulfur metabolic process	1.5	-1.4
GO:0016144	BP	2.36E-02	S-glycoside biosynthetic process	1.5	-1.4
GO:0019758	BP	2.36E-02	glycosinolate biosynthetic process	1.5	-1.4
GO:0019761	BP	2.36E-02	glucosinolate biosynthetic process	1.5	-1.4
GO:0043255	BP	2.88E-02	regulation of carbohydrate biosynthetic process	1.5	-1.4
GO:0016051	BP	3.98E-02	carbohydrate biosynthetic process	1.5	-1.4
GO:0043455	BP	4.13E-02	regulation of secondary metabolic process	1.5	-1.4
GO:0004096	MF	5.38E-03	catalase activity	1.5	-1.4
GO:0016491	MF	4.75E-02	oxidoreductase activity	1.5	-1.4
GO:0042788	CC	7.62E-03	polysomal ribosome	1.5	-0.7
GO:0005844	CC	9.96E-03	polysome	1.5	-0.7
GO:0052837	BP	1.37E-03	thiazole biosynthetic process	1.5	0.0
GO:0052838	BP	1.37E-03	thiazole metabolic process	1.5	0.0
GO:0018131	BP	4.11E-03	oxazole or thiazole biosynthetic process	1.5	0.0
GO:0046484	BP	4.11E-03	oxazole or thiazole metabolic process	1.5	0.0
GO:0006849	BP	2.05E-02	plasma membrane pyruvate transport	1.5	0.0
GO:1901475	BP	2.05E-02	pyruvate transmembrane transport	1.5	0.0
GO:0009228	BP	2.86E-02	thiamine biosynthetic process	1.5	0.0
GO:0042724	BP	2.86E-02	thiamine-containing compound biosynthetic process	1.5	0.0
GO:0006772	BP	4.89E-02	thiamine metabolic process	1.5	0.0
GO:0042723	BP	4.89E-02	thiamine-containing compound metabolic process	1.5	0.0
GO:0050833	MF	1.04E-02	pyruvate transmembrane transporter activity	1.5	0.0
GO:0016899	MF	4.97E-02	oxidoreductase activity, acting on the CH-OH group of donors, oxygen as acceptor	1.5	0.0
GO:0044434	CC	2.82E-03	chloroplast part	1.5	0.0
GO:0044435	CC	3.47E-03	plastid part	1.5	0.0
GO:0009536	CC	1.42E-02	plastid	1.5	0.0
GO:0009570	CC	2.43E-02	chloroplast stroma	1.5	0.0

GO:0009532	CC	2.82E-02	plastid stroma	1.5	0.0
GO:0009507	CC	3.00E-02	chloroplast	1.5	0.0
GO:0030599	MF	4.32E-02	pectinesterase activity	1.5	0.9
GO:0045330	MF	4.32E-02	aspartyl esterase activity	1.5	0.9
GO:0017111	MF	8.91E-04	nucleoside-triphosphatase activity	1.5	1.7
GO:0016462	MF	1.29E-03	pyrophosphatase activity	1.5	1.7
GO:0016818	MF	1.36E-03	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	1.5	1.7
GO:0016817	MF	1.49E-03	hydrolase activity, acting on acid anhydrides	1.5	1.7
GO:0003777	MF	1.79E-03	microtubule motor activity	1.5	1.7
GO:0003774	MF	4.50E-03	motor activity	1.5	1.7
GO:0016787	MF	5.58E-03	hydrolase activity	1.5	1.7
GO:0008017	MF	2.49E-02	microtubule binding	1.5	1.7
GO:0015631	MF	3.08E-02	tubulin binding	1.5	1.7
GO:0009941	CC	1.58E-03	chloroplast envelope	1.5	1.7
GO:0009526	CC	1.98E-03	plastid envelope	1.5	1.7
GO:0009570	CC	2.54E-03	chloroplast stroma	1.5	1.7
GO:0009532	CC	2.89E-03	plastid stroma	1.5	1.7
GO:0031967	CC	1.46E-02	organelle envelope	1.5	1.7
GO:0031975	CC	1.47E-02	envelope	1.5	1.7
GO:0044446	CC	4.99E-02	intracellular organelle part	1.5	1.7
GO:0007018	BP	4.14E-03	microtubule-based movement	1.5	1.7
GO:0007017	BP	5.01E-03	microtubule-based process	1.5	1.7
GO:0006928	BP	7.63E-03	movement of cell or subcellular component	1.5	1.7
GO:0003723	MF	8.67E-04	RNA binding	1.5	2.4
GO:0003729	MF	1.56E-03	mRNA binding	1.5	2.4
GO:1901363	MF	1.09E-02	heterocyclic compound binding	1.5	2.4
GO:0097159	MF	1.12E-02	organic cyclic compound binding	1.5	2.4
GO:0005200	MF	1.17E-02	structural constituent of cytoskeleton	1.5	2.4
GO:0019901	MF	1.74E-02	protein kinase binding	1.5	2.4
GO:0016273	MF	2.71E-02	arginine N-methyltransferase activity	1.5	2.4
GO:0016274	MF	2.71E-02	protein-arginine N-methyltransferase activity	1.5	2.4
GO:0019900	MF	3.46E-02	kinase binding	1.5	2.4
GO:0043228	CC	1.47E-02	non-membrane-bounded organelle	1.5	2.4
GO:0043232	CC	1.47E-02	intracellular non-membrane-bounded organelle	1.5	2.4
GO:0035145	CC	2.11E-02	exon-exon junction complex	1.5	2.4
GO:0032991	CC	3.18E-02	macromolecular complex	1.5	2.4
GO:0009987	BP	1.84E-02	cellular process	1.5	2.4
GO:0009735	BP	3.10E-02	response to cytokinin	1.5	2.4
GO:0009521	CC	3.43E-03	photosystem	1.5	3.2
GO:0009522	CC	1.43E-02	photosystem I	1.5	3.2

GO:0009538	CC	1.78E-02	photosystem I reaction center	1.5	3.2
GO:0042555	CC	2.10E-02	MCM complex	1.5	3.2
GO:0000347	CC	2.81E-02	THO complex	1.5	3.2
GO:0009579	CC	2.85E-02	thylakoid	1.5	3.2
GO:0043234	CC	3.89E-02	protein complex	1.5	3.2
GO:0000254	MF	5.60E-03	C-4 methylsterol oxidase activity	1.5	3.2
GO:0005507	MF	1.63E-02	copper ion binding	1.5	4.0
GO:0017113	MF	1.77E-02	dihydropyrimidine dehydrogenase (NADP+) activity	2.0	-4.0
GO:0102043	MF	1.80E-02	isopentenyl phosphate kinase activity	2.0	-3.0
GO:1990136	MF	4.32E-02	linoleate 9S-lipoxygenase activity	2.0	-1.3
GO:0005996	BP	5.04E-04	monosaccharide metabolic process	2.0	-0.4
GO:0080036	BP	9.76E-04	regulation of cytokinin-activated signaling pathway	2.0	-0.4
GO:0071495	BP	4.32E-03	cellular response to endogenous stimulus	2.0	-0.4
GO:0071310	BP	1.22E-02	cellular response to organic substance	2.0	-0.4
GO:0009736	BP	1.83E-02	cytokinin-activated signaling pathway	2.0	-0.4
GO:0071368	BP	1.92E-02	cellular response to cytokinin stimulus	2.0	-0.4
GO:0070887	BP	2.80E-02	cellular response to chemical stimulus	2.0	-0.4
GO:0010184	BP	3.15E-02	cytokinin transport	2.0	-0.4
GO:0051761	BP	3.76E-02	sesquiterpene metabolic process	2.0	-0.4
GO:0051762	BP	3.76E-02	sesquiterpene biosynthetic process	2.0	-0.4
GO:0032870	BP	4.67E-02	cellular response to hormone stimulus	2.0	-0.4
GO:0047918	MF	4.43E-05	GDP-mannose 3,5-epimerase activity	2.0	-0.4
GO:0051287	MF	4.48E-03	NAD binding	2.0	-0.4
GO:0080016	MF	1.01E-02	(-)-E-beta-caryophyllene synthase activity	2.0	-0.4
GO:0080017	MF	1.01E-02	alpha-humulene synthase activity	2.0	-0.4
GO:0010334	MF	1.11E-02	sesquiterpene synthase activity	2.0	-0.4
GO:0015211	MF	1.31E-02	purine nucleoside transmembrane transporter activity	2.0	-0.4
GO:0005345	MF	1.77E-02	purine nucleobase transmembrane transporter activity	2.0	-0.4
GO:0015205	MF	1.90E-02	nucleobase transmembrane transporter activity	2.0	-0.4
GO:0016857	MF	2.16E-02	racemase and epimerase activity, acting on carbohydrates and derivatives	2.0	-0.4
GO:0005337	MF	2.30E-02	nucleoside transmembrane transporter activity	2.0	-0.4
GO:0010333	MF	3.91E-02	terpene synthase activity	2.0	-0.4
GO:0016854	MF	4.28E-02	racemase and epimerase activity	2.0	-0.4
GO:1901363	MF	1.36E-02	heterocyclic compound binding	2.0	1.3
GO:0097159	MF	1.39E-02	organic cyclic compound binding	2.0	1.3
GO:0051082	MF	2.71E-02	unfolded protein binding	2.0	1.3
GO:0009846	BP	4.71E-02	pollen germination	2.0	4.0
GO:0005887	CC	1.85E-02	integral component of plasma membrane	2.5	-4.0

GO:0042026	BP	1.57E-02	protein refolding	2.5	-3.1
GO:0030245	BP	2.68E-02	cellulose catabolic process	2.5	-3.1
GO:0051275	BP	2.68E-02	beta-glucan catabolic process	2.5	-3.1
GO:0008810	MF	7.06E-03	cellulase activity	2.5	-3.1
GO:0004642	MF	1.21E-05	phosphoribosylformylglycinamidine synthase activity	2.5	-2.2
GO:0016884	MF	2.52E-03	carbon-nitrogen ligase activity, with glutamine as amido-N-donor	2.5	-2.2
GO:0016719	MF	3.19E-02	carotene 7,8-desaturase activity	2.5	-2.2
GO:0052886	MF	3.19E-02	9,9'-dicis-carotene:quinone oxidoreductase activity	2.5	-2.2
GO:0052887	MF	3.19E-02	7,9,9'-tricis-neurosporene:quinone oxidoreductase activity	2.5	-2.2
GO:0009126	BP	5.18E-04	purine nucleoside monophosphate metabolic process	2.5	-2.2
GO:0009167	BP	5.18E-04	purine ribonucleoside monophosphate metabolic process	2.5	-2.2
GO:0009161	BP	6.79E-04	ribonucleoside monophosphate metabolic process	2.5	-2.2
GO:0009123	BP	7.40E-04	nucleoside monophosphate metabolic process	2.5	-2.2
GO:0009150	BP	1.01E-03	purine ribonucleotide metabolic process	2.5	-2.2
GO:0006163	BP	1.21E-03	purine nucleotide metabolic process	2.5	-2.2
GO:0009259	BP	1.34E-03	ribonucleotide metabolic process	2.5	-2.2
GO:0009127	BP	1.64E-03	purine nucleoside monophosphate biosynthetic process	2.5	-2.2
GO:0009168	BP	1.64E-03	purine ribonucleoside monophosphate biosynthetic process	2.5	-2.2
GO:0019693	BP	2.17E-03	ribose phosphate metabolic process	2.5	-2.2
GO:0044711	BP	2.32E-03	single-organism biosynthetic process	2.5	-2.2
GO:0009156	BP	2.51E-03	ribonucleoside monophosphate biosynthetic process	2.5	-2.2
GO:0072521	BP	2.72E-03	purine-containing compound metabolic process	2.5	-2.2
GO:0009124	BP	2.95E-03	nucleoside monophosphate biosynthetic process	2.5	-2.2
GO:0009152	BP	3.85E-03	purine ribonucleotide biosynthetic process	2.5	-2.2
GO:0006189	BP	4.78E-03	'de novo' IMP biosynthetic process	2.5	-2.2
GO:0006164	BP	4.80E-03	purine nucleotide biosynthetic process	2.5	-2.2
GO:0044710	BP	5.24E-03	single-organism metabolic process	2.5	-2.2
GO:0009117	BP	5.39E-03	nucleotide metabolic process	2.5	-2.2
GO:0006753	BP	5.79E-03	nucleoside phosphate metabolic process	2.5	-2.2
GO:0009260	BP	6.19E-03	ribonucleotide biosynthetic process	2.5	-2.2
GO:0046390	BP	6.19E-03	ribose phosphate biosynthetic process	2.5	-2.2
GO:0006188	BP	6.43E-03	IMP biosynthetic process	2.5	-2.2
GO:0046040	BP	6.43E-03	IMP metabolic process	2.5	-2.2
GO:0072522	BP	8.35E-03	purine-containing compound biosynthetic process	2.5	-2.2
GO:0055086	BP	1.26E-02	nucleobase-containing small molecule metabolic process	2.5	-2.2

GO:0009165	BP	1.71E-02	nucleotide biosynthetic process	2.5	-2.2
GO:1901293	BP	1.86E-02	nucleoside phosphate biosynthetic process	2.5	-2.2
GO:0046148	BP	4.54E-02	pigment biosynthetic process	2.5	-2.2
GO:0044281	BP	4.76E-02	small molecule metabolic process	2.5	-2.2
GO:0019637	BP	4.98E-02	organophosphate metabolic process	2.5	-2.2
GO:0009507	CC	2.54E-03	chloroplast	2.5	-2.2
GO:0009536	CC	3.57E-03	plastid	2.5	-2.2
GO:0044434	CC	6.04E-03	chloroplast part	2.5	-2.2
GO:0044446	CC	6.75E-03	intracellular organelle part	2.5	-2.2
GO:0044422	CC	6.79E-03	organelle part	2.5	-2.2
GO:0044435	CC	6.81E-03	plastid part	2.5	-2.2
GO:0031967	CC	2.14E-02	organelle envelope	2.5	-2.2
GO:0031975	CC	2.15E-02	envelope	2.5	-2.2
GO:0005618	CC	1.62E-04	cell wall	2.5	-1.3
GO:0030312	CC	1.62E-04	external encapsulating structure	2.5	-1.3
GO:0044446	CC	4.32E-04	intracellular organelle part	2.5	-1.3
GO:0044422	CC	4.36E-04	organelle part	2.5	-1.3
GO:0071944	CC	6.40E-04	cell periphery	2.5	-1.3
GO:0009507	CC	9.61E-04	chloroplast	2.5	-1.3
GO:0009536	CC	1.54E-03	plastid	2.5	-1.3
GO:0005739	CC	3.18E-03	mitochondrion	2.5	-1.3
GO:0044434	CC	9.63E-03	chloroplast part	2.5	-1.3
GO:0044435	CC	1.12E-02	plastid part	2.5	-1.3
GO:0005750	CC	1.83E-02	mitochondrial respiratory chain complex III	2.5	-1.3
GO:0045275	CC	1.83E-02	respiratory chain complex III	2.5	-1.3
GO:0009570	CC	1.93E-02	chloroplast stroma	2.5	-1.3
GO:0044444	CC	2.07E-02	cytoplasmic part	2.5	-1.3
GO:0005774	CC	2.08E-02	vacuolar membrane	2.5	-1.3
GO:0044437	CC	2.11E-02	vacuolar part	2.5	-1.3
GO:1902495	CC	2.15E-02	transmembrane transporter complex	2.5	-1.3
GO:1990351	CC	2.15E-02	transporter complex	2.5	-1.3
GO:0009532	CC	2.18E-02	plastid stroma	2.5	-1.3
GO:0044464	CC	4.57E-02	cell part	2.5	-1.3
GO:0070069	CC	4.86E-02	cytochrome complex	2.5	-1.3
GO:0009678	MF	2.22E-03	hydrogen-translocating pyrophosphatase activity	2.5	-1.3
GO:0004427	MF	1.54E-02	inorganic diphosphatase activity	2.5	-1.3
GO:0010248	BP	5.75E-04	establishment or maintenance of transmembrane electrochemical gradient	2.5	-1.3
GO:1901566	BP	3.42E-03	organonitrogen compound biosynthetic process	2.5	-1.3
GO:1901605	BP	4.70E-03	alpha-amino acid metabolic process	2.5	-1.3
GO:0044281	BP	8.28E-03	small molecule metabolic process	2.5	-1.3

GO:1901607	BP	1.42E-02	alpha-amino acid biosynthetic process	2.5	-1.3
GO:0019752	BP	2.39E-02	carboxylic acid metabolic process	2.5	-1.3
GO:0008652	BP	2.51E-02	cellular amino acid biosynthetic process	2.5	-1.3
GO:0006520	BP	3.96E-02	cellular amino acid metabolic process	2.5	-1.3
GO:0019752	BP	5.16E-03	carboxylic acid metabolic process	2.5	-0.4
GO:0044281	BP	1.03E-02	small molecule metabolic process	2.5	-0.4
GO:0043436	BP	1.20E-02	oxoacid metabolic process	2.5	-0.4
GO:0006082	BP	1.22E-02	organic acid metabolic process	2.5	-0.4
GO:0043167	MF	1.67E-02	ion binding	2.5	-0.4
GO:1901001	BP	2.32E-02	negative regulation of response to salt stress	2.5	0.4
GO:0043023	MF	6.25E-04	ribosomal large subunit binding	2.5	0.4
GO:0043022	MF	7.46E-03	ribosome binding	2.5	0.4
GO:0043021	MF	2.28E-02	ribonucleoprotein complex binding	2.5	0.4
GO:0043424	MF	2.28E-02	protein histidine kinase binding	2.5	0.4
GO:0043495	MF	3.76E-02	protein anchor	2.5	1.3
GO:0003917	MF	4.83E-02	DNA topoisomerase type I activity	2.5	1.3
GO:0009908	BP	1.33E-03	flower development	2.5	2.2
GO:0090567	BP	1.50E-03	reproductive shoot system development	2.5	2.2
GO:0010077	BP	2.21E-03	maintenance of inflorescence meristem identity	2.5	2.2
GO:0009791	BP	1.37E-02	post-embryonic development	2.5	2.2
GO:0048367	BP	1.58E-02	shoot system development	2.5	2.2
GO:0006355	BP	1.65E-02	regulation of transcription, DNA-templated	2.5	2.2
GO:1903506	BP	1.65E-02	regulation of nucleic acid-templated transcription	2.5	2.2
GO:2001141	BP	1.66E-02	regulation of RNA biosynthetic process	2.5	2.2
GO:0051252	BP	1.83E-02	regulation of RNA metabolic process	2.5	2.2
GO:0019219	BP	2.23E-02	regulation of nucleobase-containing compound metabolic process	2.5	2.2
GO:0006351	BP	2.50E-02	transcription, DNA-templated	2.5	2.2
GO:0097659	BP	2.50E-02	nucleic acid-templated transcription	2.5	2.2
GO:0032774	BP	2.52E-02	RNA biosynthetic process	2.5	2.2
GO:2000112	BP	2.52E-02	regulation of cellular macromolecule biosynthetic process	2.5	2.2
GO:0010556	BP	2.57E-02	regulation of macromolecule biosynthetic process	2.5	2.2
GO:0010582	BP	2.72E-02	floral meristem determinacy	2.5	2.2
GO:0010022	BP	3.07E-02	meristem determinacy	2.5	2.2
GO:0048731	BP	3.15E-02	system development	2.5	2.2
GO:0031326	BP	3.24E-02	regulation of cellular biosynthetic process	2.5	2.2
GO:0009889	BP	3.46E-02	regulation of biosynthetic process	2.5	2.2
GO:0034654	BP	3.94E-02	nucleobase-containing compound biosynthetic process	2.5	2.2
GO:0010468	BP	4.00E-02	regulation of gene expression	2.5	2.2

GO:0051171	BP	4.47E-02	regulation of nitrogen compound metabolic process	2.5	2.2
GO:0010074	BP	4.85E-02	maintenance of meristem identity	2.5	2.2
GO:0005634	CC	6.41E-03	nucleus	2.5	2.2
GO:0003700	MF	8.21E-06	transcription factor activity, sequence-specific DNA binding	2.5	2.2
GO:0001071	MF	8.43E-06	nucleic acid binding transcription factor activity	2.5	2.2
GO:0042802	MF	9.16E-06	identical protein binding	2.5	2.2
GO:0003677	MF	1.85E-03	DNA binding	2.5	2.2
GO:0031490	MF	1.85E-03	chromatin DNA binding	2.5	2.2
GO:0042803	MF	5.20E-03	protein homodimerization activity	2.5	2.2
GO:0043621	MF	2.29E-02	protein self-association	2.5	2.2
GO:0033925	MF	2.61E-02	mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase activity	2.5	2.2
GO:0003682	MF	3.31E-02	chromatin binding	2.5	2.2
GO:0003676	MF	4.03E-02	nucleic acid binding	2.5	2.2
GO:0016307	MF	3.14E-03	phosphatidylinositol phosphate kinase activity	2.5	3.1
GO:0008652	BP	4.10E-02	cellular amino acid biosynthetic process	3.0	-4.0
GO:0003968	MF	3.95E-06	RNA-directed 5'-3' RNA polymerase activity	3.0	-3.1
GO:0034062	MF	2.79E-03	5'-3' RNA polymerase activity	3.0	-3.1
GO:0097747	MF	2.79E-03	RNA polymerase activity	3.0	-3.1
GO:0016779	MF	2.90E-02	nucleotidyltransferase activity	3.0	-3.1
GO:0010495	BP	2.78E-06	long-distance posttranscriptional gene silencing	3.0	-3.1
GO:0060148	BP	4.17E-06	positive regulation of posttranscriptional gene silencing	3.0	-3.1
GO:0070919	BP	1.81E-05	production of siRNA involved in chromatin silencing by small RNA	3.0	-3.1
GO:0060147	BP	6.58E-05	regulation of posttranscriptional gene silencing	3.0	-3.1
GO:0031048	BP	1.44E-04	chromatin silencing by small RNA	3.0	-3.1
GO:0042221	BP	5.70E-04	response to chemical	3.0	-3.1
GO:0030422	BP	1.14E-03	production of siRNA involved in RNA interference	3.0	-3.1
GO:0060968	BP	1.28E-03	regulation of gene silencing	3.0	-3.1
GO:0016246	BP	1.43E-03	RNA interference	3.0	-3.1
GO:0010025	BP	1.75E-03	wax biosynthetic process	3.0	-3.1
GO:1901570	BP	1.84E-03	fatty acid derivative biosynthetic process	3.0	-3.1
GO:0010166	BP	1.93E-03	wax metabolic process	3.0	-3.1
GO:1901568	BP	2.43E-03	fatty acid derivative metabolic process	3.0	-3.1
GO:0031050	BP	4.27E-03	dsRNA fragmentation	3.0	-3.1
GO:0043331	BP	4.27E-03	response to dsRNA	3.0	-3.1
GO:0070918	BP	4.27E-03	production of small RNA involved in gene silencing by RNA	3.0	-3.1
GO:0071359	BP	4.27E-03	cellular response to dsRNA	3.0	-3.1
GO:0010033	BP	5.57E-03	response to organic substance	3.0	-3.1

GO:0006342	BP	5.65E-03	chromatin silencing	3.0	-3.1
GO:0045814	BP	6.85E-03	negative regulation of gene expression, epigenetic	3.0	-3.1
GO:0035194	BP	7.74E-03	posttranscriptional gene silencing by RNA	3.0	-3.1
GO:1901700	BP	1.35E-02	response to oxygen-containing compound	3.0	-3.1
GO:0016441	BP	1.47E-02	posttranscriptional gene silencing	3.0	-3.1
GO:0044711	BP	1.70E-02	single-organism biosynthetic process	3.0	-3.1
GO:0009615	BP	2.39E-02	response to virus	3.0	-3.1
GO:1901698	BP	3.84E-02	response to nitrogen compound	3.0	-3.1
GO:1901699	BP	4.95E-02	cellular response to nitrogen compound	3.0	-3.1
GO:0016866	MF	6.94E-04	intramolecular transferase activity	3.0	-2.2
GO:0003735	MF	2.74E-02	structural constituent of ribosome	3.0	-2.2
GO:0022626	CC	7.69E-04	cytosolic ribosome	3.0	-2.2
GO:0044445	CC	8.62E-04	cytosolic part	3.0	-2.2
GO:0005840	CC	5.38E-03	ribosome	3.0	-2.2
GO:0030529	CC	1.71E-02	intracellular ribonucleoprotein complex	3.0	-2.2
GO:1990904	CC	1.71E-02	ribonucleoprotein complex	3.0	-2.2
GO:0030244	BP	5.43E-03	cellulose biosynthetic process	3.0	-2.2
GO:0051274	BP	1.02E-02	beta-glucan biosynthetic process	3.0	-2.2
GO:0030243	BP	1.52E-02	cellulose metabolic process	3.0	-2.2
GO:0051273	BP	2.39E-02	beta-glucan metabolic process	3.0	-2.2
GO:0009250	BP	4.97E-02	glucan biosynthetic process	3.0	-2.2
GO:0000160	BP	4.79E-03	phosphorelay signal transduction system	3.0	-1.3
GO:0043231	CC	7.08E-03	intracellular membrane-bounded organelle	3.0	-0.4
GO:0043227	CC	8.07E-03	membrane-bounded organelle	3.0	-0.4
GO:0043229	CC	1.44E-02	intracellular organelle	3.0	-0.4
GO:0043226	CC	1.45E-02	organelle	3.0	-0.4
GO:0044391	CC	4.12E-48	ribosomal subunit	3.5	-4.0
GO:0005840	CC	4.28E-43	ribosome	3.5	-4.0
GO:0030529	CC	1.03E-39	intracellular ribonucleoprotein complex	3.5	-4.0
GO:1990904	CC	1.03E-39	ribonucleoprotein complex	3.5	-4.0
GO:0022626	CC	1.81E-39	cytosolic ribosome	3.5	-4.0
GO:0044445	CC	4.52E-39	cytosolic part	3.5	-4.0
GO:0015935	CC	9.32E-37	small ribosomal subunit	3.5	-4.0
GO:0022627	CC	7.36E-33	cytosolic small ribosomal subunit	3.5	-4.0
GO:0043228	CC	2.04E-30	non-membrane-bounded organelle	3.5	-4.0
GO:0043232	CC	2.04E-30	intracellular non-membrane-bounded organelle	3.5	-4.0
GO:0032991	CC	1.35E-22	macromolecular complex	3.5	-4.0
GO:0005730	CC	1.11E-18	nucleolus	3.5	-4.0
GO:0005829	CC	3.77E-16	cytosol	3.5	-4.0
GO:0071944	CC	2.05E-14	cell periphery	3.5	-4.0
GO:0031981	CC	2.86E-14	nuclear lumen	3.5	-4.0

GO:0031974	CC	1.73E-13	membrane-enclosed lumen	3.5	-4.0
GO:0043233	CC	1.73E-13	organelle lumen	3.5	-4.0
GO:0070013	CC	1.73E-13	intracellular organelle lumen	3.5	-4.0
GO:0042788	CC	2.83E-13	polysomal ribosome	3.5	-4.0
GO:0022625	CC	3.51E-13	cytosolic large ribosomal subunit	3.5	-4.0
GO:0005844	CC	1.30E-12	polysome	3.5	-4.0
GO:0015934	CC	1.46E-12	large ribosomal subunit	3.5	-4.0
GO:0044428	CC	3.29E-12	nuclear part	3.5	-4.0
GO:0005618	CC	6.74E-11	cell wall	3.5	-4.0
GO:0030312	CC	6.74E-11	external encapsulating structure	3.5	-4.0
GO:0044446	CC	3.70E-10	intracellular organelle part	3.5	-4.0
GO:0044422	CC	3.79E-10	organelle part	3.5	-4.0
GO:0043229	CC	4.98E-10	intracellular organelle	3.5	-4.0
GO:0043226	CC	5.14E-10	organelle	3.5	-4.0
GO:0009506	CC	6.36E-09	plasmodesma	3.5	-4.0
GO:0055044	CC	6.36E-09	symplast	3.5	-4.0
GO:0005911	CC	6.50E-09	cell-cell junction	3.5	-4.0
GO:0030054	CC	6.50E-09	cell junction	3.5	-4.0
GO:0005737	CC	7.59E-09	cytoplasm	3.5	-4.0
GO:0044444	CC	1.48E-08	cytoplasmic part	3.5	-4.0
GO:0044464	CC	1.50E-08	cell part	3.5	-4.0
GO:0005886	CC	1.67E-08	plasma membrane	3.5	-4.0
GO:0005623	CC	2.21E-08	cell	3.5	-4.0
GO:0044424	CC	7.90E-08	intracellular part	3.5	-4.0
GO:0005622	CC	1.87E-07	intracellular	3.5	-4.0
GO:0009507	CC	1.88E-07	chloroplast	3.5	-4.0
GO:0005773	CC	3.40E-07	vacuole	3.5	-4.0
GO:0009536	CC	6.24E-07	plastid	3.5	-4.0
GO:0005774	CC	1.46E-06	vacuolar membrane	3.5	-4.0
GO:0044437	CC	1.51E-06	vacuolar part	3.5	-4.0
GO:0005634	CC	8.76E-06	nucleus	3.5	-4.0
GO:0098805	CC	3.21E-05	whole membrane	3.5	-4.0
GO:0043231	CC	5.39E-05	intracellular membrane-bounded organelle	3.5	-4.0
GO:0043227	CC	8.07E-05	membrane-bounded organelle	3.5	-4.0
GO:0098588	CC	1.20E-04	bounding membrane of organelle	3.5	-4.0
GO:0031090	CC	1.88E-03	organelle membrane	3.5	-4.0
GO:0016020	CC	3.95E-03	membrane	3.5	-4.0
GO:0006412	BP	6.57E-34	translation	3.5	-4.0
GO:0043043	BP	1.54E-33	peptide biosynthetic process	3.5	-4.0
GO:0006518	BP	6.61E-33	peptide metabolic process	3.5	-4.0
GO:0043604	BP	9.70E-33	amide biosynthetic process	3.5	-4.0
GO:0043603	BP	1.38E-31	cellular amide metabolic process	3.5	-4.0

GO:1901566	BP	5.17E-24	organonitrogen compound biosynthetic process	3.5	-4.0
GO:0042256	BP	4.34E-20	mature ribosome assembly	3.5	-4.0
GO:0042255	BP	4.09E-19	ribosome assembly	3.5	-4.0
GO:0070925	BP	1.55E-17	organelle assembly	3.5	-4.0
GO:0022618	BP	6.85E-16	ribonucleoprotein complex assembly	3.5	-4.0
GO:0071826	BP	6.85E-16	ribonucleoprotein complex subunit organization	3.5	-4.0
GO:0042254	BP	7.70E-16	ribosome biogenesis	3.5	-4.0
GO:1901576	BP	2.29E-14	organic substance biosynthetic process	3.5	-4.0
GO:0022613	BP	4.10E-14	ribonucleoprotein complex biogenesis	3.5	-4.0
GO:0044249	BP	6.57E-14	cellular biosynthetic process	3.5	-4.0
GO:0034645	BP	8.04E-14	cellular macromolecule biosynthetic process	3.5	-4.0
GO:0044271	BP	9.95E-14	cellular nitrogen compound biosynthetic process	3.5	-4.0
GO:0009058	BP	1.25E-13	biosynthetic process	3.5	-4.0
GO:0009059	BP	1.59E-13	macromolecule biosynthetic process	3.5	-4.0
GO:0044267	BP	6.77E-13	cellular protein metabolic process	3.5	-4.0
GO:0010467	BP	1.92E-12	gene expression	3.5	-4.0
GO:0019538	BP	6.69E-11	protein metabolic process	3.5	-4.0
GO:0034622	BP	1.22E-09	cellular macromolecular complex assembly	3.5	-4.0
GO:0065003	BP	1.61E-08	macromolecular complex assembly	3.5	-4.0
GO:0034641	BP	2.08E-08	cellular nitrogen compound metabolic process	3.5	-4.0
GO:1901564	BP	2.53E-08	organonitrogen compound metabolic process	3.5	-4.0
GO:0043933	BP	4.97E-08	macromolecular complex subunit organization	3.5	-4.0
GO:0044237	BP	1.75E-07	cellular metabolic process	3.5	-4.0
GO:0022607	BP	5.76E-07	cellular component assembly	3.5	-4.0
GO:0044085	BP	4.53E-06	cellular component biogenesis	3.5	-4.0
GO:0044260	BP	4.59E-06	cellular macromolecule metabolic process	3.5	-4.0
GO:0071704	BP	5.00E-06	organic substance metabolic process	3.5	-4.0
GO:0009987	BP	7.09E-06	cellular process	3.5	-4.0
GO:0008152	BP	7.62E-06	metabolic process	3.5	-4.0
GO:0044238	BP	9.78E-06	primary metabolic process	3.5	-4.0
GO:0006996	BP	4.97E-05	organelle organization	3.5	-4.0
GO:0043170	BP	1.74E-04	macromolecule metabolic process	3.5	-4.0
GO:0006807	BP	7.70E-04	nitrogen compound metabolic process	3.5	-4.0
GO:0009735	BP	8.03E-04	response to cytokinin	3.5	-4.0
GO:0071731	BP	6.48E-03	response to nitric oxide	3.5	-4.0
GO:0009965	BP	7.16E-03	leaf morphogenesis	3.5	-4.0
GO:0071840	BP	2.40E-02	cellular component organization or biogenesis	3.5	-4.0
GO:0009725	BP	3.99E-02	response to hormone	3.5	-4.0
GO:0009719	BP	4.15E-02	response to endogenous stimulus	3.5	-4.0

GO:0003735	MF	3.66E-45	structural constituent of ribosome	3.5	-4.0
GO:0005198	MF	7.65E-42	structural molecule activity	3.5	-4.0
GO:0003723	MF	2.71E-19	RNA binding	3.5	-4.0
GO:0003729	MF	7.66E-17	mRNA binding	3.5	-4.0
GO:0003676	MF	2.51E-10	nucleic acid binding	3.5	-4.0
GO:0019843	MF	9.30E-08	rRNA binding	3.5	-4.0
GO:1901363	MF	1.90E-05	heterocyclic compound binding	3.5	-4.0
GO:0097159	MF	1.99E-05	organic cyclic compound binding	3.5	-4.0
GO:0016702	MF	4.93E-03	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	3.5	-4.0
GO:0016701	MF	1.63E-02	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	3.5	-4.0
GO:0051213	MF	1.91E-02	dioxygenase activity	3.5	-4.0
GO:0048467	BP	4.42E-03	gynoecium development	3.5	-3.1
GO:0042254	BP	3.43E-02	ribosome biogenesis	3.5	-3.1
GO:0000977	MF	3.16E-03	RNA polymerase II regulatory region sequence-specific DNA binding	3.5	-2.2
GO:0001012	MF	3.16E-03	RNA polymerase II regulatory region DNA binding	3.5	-2.2
GO:2001066	MF	1.53E-02	amylopectin binding	3.5	-2.2
GO:0000976	MF	2.07E-02	transcription regulatory region sequence-specific DNA binding	3.5	-2.2
GO:0004721	MF	3.00E-02	phosphoprotein phosphatase activity	3.5	-2.2
GO:1990837	MF	3.20E-02	sequence-specific double-stranded DNA binding	3.5	-2.2
GO:0009911	BP	2.80E-02	positive regulation of flower development	3.5	-2.2
GO:0045944	BP	2.93E-02	positive regulation of transcription from RNA polymerase II promoter	3.5	-2.2
GO:0003690	MF	4.43E-03	double-stranded DNA binding	3.5	-0.4
GO:0000977	MF	7.49E-03	RNA polymerase II regulatory region sequence-specific DNA binding	3.5	-0.4
GO:0001012	MF	7.49E-03	RNA polymerase II regulatory region DNA binding	3.5	-0.4
GO:0046983	MF	9.08E-03	protein dimerization activity	3.5	-0.4
GO:0010313	MF	2.45E-02	phytochrome binding	3.5	-0.4
GO:0010114	BP	1.16E-02	response to red light	3.5	-0.4
GO:0048440	BP	1.37E-02	carpel development	3.5	-0.4
GO:0080112	BP	1.88E-02	seed growth	3.5	-0.4
GO:0048438	BP	2.46E-02	floral whorl development	3.5	-0.4
GO:0048467	BP	2.69E-02	gynoecium development	3.5	-0.4
GO:0009900	BP	4.17E-02	dehiscence	3.5	-0.4
GO:0003949	MF	1.37E-02	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase activity	3.5	0.4
GO:0004793	MF	1.37E-02	threonine aldolase activity	3.5	0.4
GO:0006520	BP	4.13E-02	cellular amino acid metabolic process	3.5	0.4

GO:0033106	CC	5.17E-03	cis-Golgi network membrane	3.5	0.4
GO:0044212	MF	3.92E-02	transcription regulatory region DNA binding	3.5	1.3
GO:0000975	MF	4.19E-02	regulatory region DNA binding	3.5	1.3
GO:0001067	MF	4.19E-02	regulatory region nucleic acid binding	3.5	1.3
GO:0009705	CC	4.23E-02	plant-type vacuole membrane	3.5	1.3
GO:0016743	MF	3.47E-02	carboxyl- or carbamoyltransferase activity	3.5	2.2
GO:0004557	MF	2.74E-03	alpha-galactosidase activity	3.5	4.0
GO:0052692	MF	2.74E-03	raffinose alpha-galactosidase activity	3.5	4.0
GO:0015925	MF	4.85E-02	galactosidase activity	3.5	4.0
GO:0044249	BP	9.89E-03	cellular biosynthetic process	4.0	-4.0
GO:1901576	BP	1.24E-02	organic substance biosynthetic process	4.0	-4.0
GO:0009058	BP	2.08E-02	biosynthetic process	4.0	-4.0
GO:0022627	CC	3.99E-03	cytosolic small ribosomal subunit	4.0	-4.0
GO:0015935	CC	7.31E-03	small ribosomal subunit	4.0	-4.0
GO:0044391	CC	2.37E-02	ribosomal subunit	4.0	-4.0
GO:0009506	CC	2.96E-02	plasmodesma	4.0	-4.0
GO:0055044	CC	2.96E-02	symplast	4.0	-4.0
GO:0005911	CC	2.98E-02	cell-cell junction	4.0	-4.0
GO:0030054	CC	2.98E-02	cell junction	4.0	-4.0
GO:0000996	MF	2.18E-02	core DNA-dependent RNA polymerase binding promoter specificity activity	4.0	-0.4
GO:0016987	MF	2.18E-02	sigma factor activity	4.0	-0.4
GO:0016805	MF	3.06E-03	dipeptidase activity	4.0	0.0
GO:0016843	MF	3.95E-02	amine-lyase activity	4.0	0.0
GO:0016844	MF	3.95E-02	strictosidine synthase activity	4.0	0.0
GO:0010494	CC	3.32E-02	cytoplasmic stress granule	4.0	0.0
GO:0017046	MF	1.57E-02	peptide hormone binding	4.0	0.4
GO:0006914	BP	2.71E-02	autophagy	4.0	0.4
GO:1905392	BP	5.53E-07	plant organ morphogenesis	4.0	3.1
GO:0099402	BP	3.94E-05	plant organ development	4.0	3.1
GO:0050896	BP	5.63E-05	response to stimulus	4.0	3.1
GO:0010015	BP	5.82E-05	root morphogenesis	4.0	3.1
GO:0048731	BP	8.54E-05	system development	4.0	3.1
GO:0044707	BP	1.95E-04	single-multicellular organism process	4.0	3.1
GO:0048364	BP	3.65E-04	root development	4.0	3.1
GO:0022622	BP	3.93E-04	root system development	4.0	3.1
GO:0007275	BP	4.78E-04	multicellular organism development	4.0	3.1
GO:0032501	BP	5.51E-04	multicellular organismal process	4.0	3.1
GO:0044699	BP	7.63E-04	single-organism process	4.0	3.1
GO:0044767	BP	9.40E-04	single-organism developmental process	4.0	3.1
GO:0009888	BP	1.23E-03	tissue development	4.0	3.1
GO:0010053	BP	1.33E-03	root epidermal cell differentiation	4.0	3.1
GO:0032502	BP	1.68E-03	developmental process	4.0	3.1

GO:0009653	BP	2.28E-03	anatomical structure morphogenesis	4.0	3.1
GO:0090627	BP	2.83E-03	plant epidermal cell differentiation	4.0	3.1
GO:0048856	BP	3.80E-03	anatomical structure development	4.0	3.1
GO:0090558	BP	4.40E-03	plant epidermis development	4.0	3.1
GO:0009965	BP	6.62E-03	leaf morphogenesis	4.0	3.1
GO:0010016	BP	7.30E-03	shoot system morphogenesis	4.0	3.1
GO:0003002	BP	8.36E-03	regionalization	4.0	3.1
GO:0048366	BP	1.01E-02	leaf development	4.0	3.1
GO:0048367	BP	1.16E-02	shoot system development	4.0	3.1
GO:0030154	BP	1.28E-02	cell differentiation	4.0	3.1
GO:0007389	BP	1.67E-02	pattern specification process	4.0	3.1
GO:0050793	BP	2.04E-02	regulation of developmental process	4.0	3.1
GO:0006950	BP	3.35E-02	response to stress	4.0	3.1
GO:0042221	BP	3.69E-02	response to chemical	4.0	3.1
GO:0090353	MF	4.73E-02	polygalacturonase inhibitor activity	4.0	3.1
GO:0009664	BP	1.02E-02	plant-type cell wall organization	4.0	4.0
GO:0071555	BP	2.56E-02	cell wall organization	4.0	4.0