Lit Review

Table of Contents:

Abbreviations 3

Gene Abbreviations 4

Safflower – History and Characteristics 6

Vernalisation 7

History of Vernalisation 8

The Biology of Vernalisation 8

Vernalisation Pathways in *Arabidopsis thaliana* 9

Vernalisation in Cereals 13

Vernalisation in Beets 15

Determining Factors Affecting Vernalisation 15

Sequencing Methodologies 16

Sanger Sequencing 16

Expressed Sequence Tags (ESTs) 17

Microarrays 18

Next Generation Sequencing (NGS) 18

Assessing the Quality of a *de novo* Assembly 18

Aims of this research project: 19

References 20

# Abbreviations

18:1 – oleic acid (double bond between the 9th and 10th carbon)

18:2 – linoleic acid (double bond between the 9th and 10thand 12th and 13th carbons)

*A. thaliana – Arabidopsis thaliana* (thale cress)

*B. vulgaris – Beta vulgaris* (beet)

bp – base pair

Col-0 – *A. thaliana* ecotype “Colombia”

DAG – diacyl glyceride

ddNTP – di-deoxynucleotide triphosphate

dNTP – deoxynucleotide triphosphate

EST – Expressed Sequence Tag

*H. vulgare – Hordeum vulgare* (barley)

HDAC – Histone deacetylase complex

*Ler*-0 – *A. thaliana* ecotype *“*Landsberg *erecta*”

NCBI - National Center for Biotechnology Information

*O. sativa – Oryza sativa* (rice)

PC – phosphatidylcholine

PCR – polymerase chain reaction

PEBP – phosphatidylethanolamine-binding protein

PHD – Plant Homeo Domain

PRC2 – Polycomb Repression Complex 2

*T. aestivum – Triticum aestivum* (bread wheat)

# Gene Abbreviations

*AP* – APETALA

*BTC1* – BOLTING TIME CONTOL 1

*CAB* – CHLOROPHYL A/B BINDING

*CAL* – CAULIFLOWER

*CCT* – CONSTANS, CONSTANS-LIKE AND TOC

CLF *–* CURLY LEAF

*CO* – CONSTANS

*FCA –* FLOWERING TIME CONTROL PROTEIN

*FIE* – FERTILIZATION-INDEPENDENT ENDOSPERM

*FIP* – FRIGIDA INTERACTING PROTEIN

*FL -*

*FLC* – FLOWERING LOCUS C

*FLD* – FLOWERING LOCUS D

*FLK* – plant specific K-homology (KH)-domain RNA-binding protein

*FPA* – FLOWERING TIME CONTROL PROTEIN

*FRI* – FRIGIDA

*FRL* – FRIGIDA-LIKE

*FT* – FLOWERING TIME

*FUL* – FRUITFUL

*FVE* – mammalian homologue of retinoblastoma-associated protein (RbAp)

*FY – A. thaliana* homologue of yeast RNA 3' processing factor (Pfs2p)

*HvCCT* – CONSTANS, CONSTANS-like and TOC (*H. vulgare)*

*LD* – LUMINODEPENDENS (or LONG DAY)

*LFY* – LEAFY

*MAF* – MADS-box AFFECTING FLOWERING

*OS2* – ODDSOC2

*PPD* – PHOTOPERIOD

*SWN* – SWINGER

*SOC1* – SUPPRESSOR OF OVEREXPRESSION OF CO1

*TOC* – TIMING OF *CAB* EXPRESSION

*TSF –* TWIN SISTER OF *FT*

*VRN1|2* – VERNALISATION (NB: *VRN1* and *VRN2* in cereals are different to *VRN1* and *VRN2* in *A. thaliana*)

*VIN3* – VERNALISATION INSENSITIVE 3

*Gene Abbreviations Note 1* – Where a gene exists as a variant in a specific species, the gene will be prefixed with a two letter abbreviation of the species i.e. *A. thaliana* (*At*), *Beta vulgaris* (*Bv*)

*Gene Abbreviations Note 2* – When a gene has a recessive allele, it will be indicated as lower case i.e. *FLOWERING TIME – FT* (dominant, normal) vs. *ft* (recessive)

# Safflower – History and Characteristics

Safflower (*Carthamus tinctorius* L.) is an Asteraceae native to Eastern and Southern Mediterranean regions, and the Middle East and India. Safflower has been cultivated in these regions for thousands of years, with safflower arrangements and safflower based dyes found in Pharaoh Tutankhamen’s tomb as well as at archaeological sites dating back to ancient Mesopotamia (Zohary & Hopf, 1993). It was originally cultivated for its edible seeds as well as dyes created from its vibrant yellow and orange flowers. Safflower seed oil is similar to sunflower seed oil (*Helianthus annuus* L.); both have similar melting points and fatty acid composition (Chempro Technoatvation Pvt. Ltd. n.d.). However, safflower oil tends to be less susceptible to oxidation, lending itself for use as a base for varnishes and oil based paints (Işigigür et al., 1995; Gecgel et al., 2007) as well as industrial lubricant applications. After oil extraction, safflower meal can be used as an animal feedstock. In 2012, just over 833,793 tonnes (t) of safflower seed was harvested globally ( Figure 1), with approximately 65% (536,651 t) originating from Mexico (30.9%, 257 451 t), 18.2% from India (152,000 t) and 15.2% from Kazakhstan (127,200 t) (United Nations 2014). Australian produced safflower seed accounted for just over 0.5% (4,800 t) of the 2012 global production, as it is more commonly grown as a ‘break crop’, to break up hard claypans or to remove excess water from soils before the cultivation of more traditional crops, such as wheat and barley (Knights, 2010).

Of particular interest is the coordination of safflower flowering time, a trait that if modifiable is likely to have impact on both safflower’s adaptability to climate change, and total yield. Research in the late 1970s characterised winter and spring varieties of safflower, specifically describing low survival of spring safflower when planted in winter and conversely; poor performance of winter safflower when planted in spring (Yazdi-Samadi & Zali, 1979). Research in *Arabidopsis thaliana* (*Arabidopsis*), and in cereals, indicates that molecular modifications to vernalisation is the primary classifier of a winter or spring ecotype within these species (Refs). Therefore, it is hypothesised, and this hypothesis is supported by previous research (Johnson et al., 2006; Refs), that “winter hardy” varieties of safflower will possess and express molecular phenotypes similar to those previously reported in other species. However, it is also important to note, that while some safflower ecotypes rapidly proceed to flowering following vernalisation, it is not a necessary environmental cue for flowering in safflower ; instead, vernalisation is seen as a *facultative* response (Salisbury & Ross, 1992). While it has been known for some time that vernalisation relates to crop yield in many important species, including wheat, the genetic mechanism(s) of vernalisation in safflower, and the effect of an extended cold treatment, remains to be characterised.

Safflower is a diploid of 11 chromosomal pairs (Knowles 2012). The approximate haploid genome size in four safflower cultivars, namely “Ljubljana”, “Uzbekistan”,   
“S-2190” and “Huesca” has been calculated as 1.34 Gbp, 1.38 Gbp, 1.39 Gbp and 1.40 Gbp respectively (Garnatje et al. 2006). However, while there are a number of online transcriptomic resources available for safflower (Li et al. 2011; Li et al. 2012; Lulin et al. 2012) and publicly accessible EST resources such as the National Center

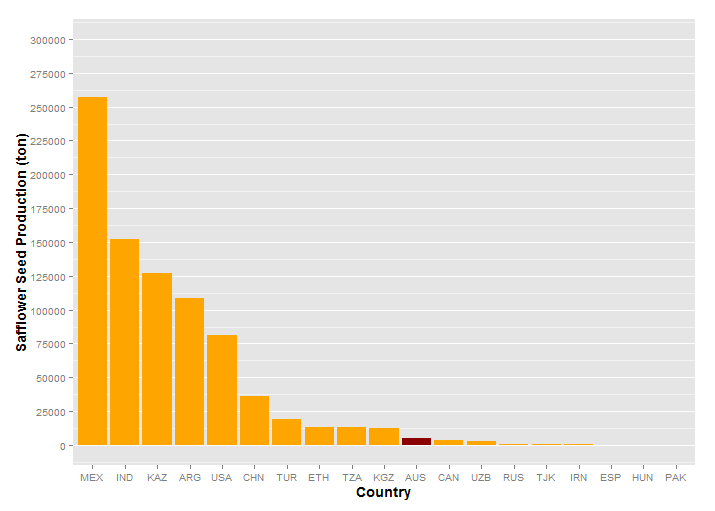


Figure 1 – Global safflower production in 2012 ordered by quantity. Australia (AUS, indicated) is ranked 11th in the world for safflower seed production at approximately 4 800 t.

for Biotechnology Information (NCBI), currently, there is no draft genome for safflower or other Asteraceae species. Furthermore, little known about the molecular evolution of Asteraceae members in one of the largest and most successful flowering plant families, especially within the context of vernalisation

# Vernalisation

Vernalisation, and its effect on harvest time and total crop yield has been a central research focus of the plant biology community for over 150 years. Vernalisation is characterised by a prolonged, greater than two weeks, exposure to low, yet non-freezing temperatures (refs). In addition, (which plants????), it has been demonstrated that the resulting time to flowering is directly proportional to the period of vernalisation (Sheldon *et al*., 2000), and furthermore, in numerous plant species planting time has been optimised to increase the period of time a plant remains in the vegetative growth stage. For example, carrots store carbohydrates in their root organ during their optimised growing season, and then following a “wintering” period will transition into reproductive growth and utilise this stored source of energy for flowering and seed production (Ingram *et al*., 2008). While flowering is often attributed to increased day length, day length is not solely responsible for the invocation of flowering. Exposure of a seedling to an extend period of cold is also responsible for triggering an early transition from vegetative to reproductive growth, and this transition is largely the result initiation of the molecular flowering pathway. Vernalisation is an important determinant of flowering time, and flowering time is an important component of overall yield for a plant. Vernalisation responses also protect delicate organs (such as those created during flowering and required for reproduction) from damage to cold exposure, restricting development until after winter has passed.

## History of Vernalisation

As early as 1857 the necessity of vernalisation for some “winter” cereals to flower was demonstrated (Klippaert, 1857). This initial demonstration was further investigated in other crop species by Gassner (1918), and the term ‘vernalisation’ was coined by Lysenko in 1928 (Latin: *vernum* meaning *spring*), who conducted a vast volume of agricultural research in the Soviet Union (reviewed by Chouard, 1960). These early studies by Lysenko and others showed that some cereal cultivars germinated as the weather warmed even though they were planted prior to, or during winter (refs). However, Lysenko incorrectly stated that the progeny of a ‘vernalised’ cereal maintained the attributes of the vernalised parent plant, and did not require re-exposure winter temperatures for germination. I addition, in many early Australian colonies, plants that had been transported from Europe struggled under the warmer conditions and mild winters of the Australian environment and this lead to widespread hunger for the early settlers as crops failed (refs). It was not until almost XXX years later through the work of William Farrar that many Australian wheats were developed via a traditional cross breeding approach using European wheats and selecting those progeny best suited for the Australian climate (Macindoe & Brown, 1968).

## The Biology of Vernalisation

Grafting was originally used to demonstrate the transmissible characteristic of vernalised shoot tissue (reviewed by Chouard, 1960). Namely, when a vernalised shoot apical meristem (SAM) was grafted on to a non-vernalised root stock, the grafted plant will flower as if the entire plant had been exposed to the vernalisation treatment. Conversely, when a non-vernalised shoot tip (or SAM – need to be specific) was grafted on to a vernalised root stock, the opposite was observed. This observation has been consistently reported for both facultative (vernalisation decreases the time period to flowering, but is not essential for flowering) and absolute (vernalisation is required for the plant to transition from vegetative to reproductive growth) vernalisation sensitive species. Contemporary research has shown that the regulatory, and genetic mechanisms of vernalisation are species-specific (Fig. 2)figure 2 below). For example,vernalisation in *A. thaliana*,and in many other dicotyledonous species, is centrally regulated by the expression of the floral repressor, *FLOWERING LOCUS C* (*FLC*)expression, whereas cereals such as

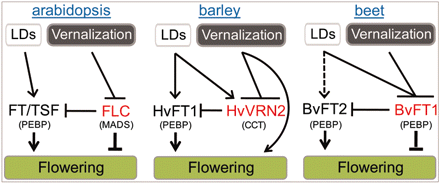


Figure 2 – The metabolic pathways central to vernalisation in *A. thaliana*, barley (*Hordeum vulgare* L.) and beet (*Beta vulgaris*) (Pin et al. 2010)

barley are regulated via *VRN2*. Beets (*Beta vulgaris* ssp. *Vulgaris*) have a different genetic mechanism, where *FT1* is responsible for regulating *FT2* (Pin et al. 2010). The commonality in each of these examples is that all changes are epigenetic in nature (i.e. environmental changes modify gene expression resulting in a different phenotype being expressed). In each of the examples shown in figure 2, regardless of species, phenotypic attributes brought on by vernalisation are epigenetic in nature. Thus, while genetically identical plants responsive to vernalisation will produce different phenotypes when vernalised, the progeny of each of these plants will produce the same vernalisation phenotype, regardless of the parental phenotype.

However, it must also be known that the above mentioned vernalisation pathways are not always seen in every cultivar of each species. In *A. thaliana*, “summer” (i.e. facultative vernalisation ecotypes) and “winter” (i.e. absolute vernalisation ecotypes) cultivars, sourced from the wild or developed for vernalisation research, and the existence of summer and winter wheats, barley and oats further demonstrates differences in the underlying genetic pathways resulting in differing phenotypes.

## Vernalisation Pathways in *Arabidopsis thaliana*

In *Arabidopsis*,FLCis a MADS-box transcription factor that mediates the transition of *Arabidopsis*, and many other dicots, from vegetative to reproductive growth (see figure 3 below). Research indicates that both genetic and epigenetic mechanisms contribute to repressing the expression of the floral repressor *FLC* during vegetative to reproductive growth transition (Boss et al. 2004; Finnegan et al. 2005). When *FLC* is expressed at high levels, promoted by FRIGIDA (FRI), FRIGIDA-LIKE1(FRL1) and FRIGIDALIKE 2 (FRL2), FLC represses the *FLOWERING TIME* (*FT*) expression, and the expression of the FT homologues, *TWIN SISTER OF*

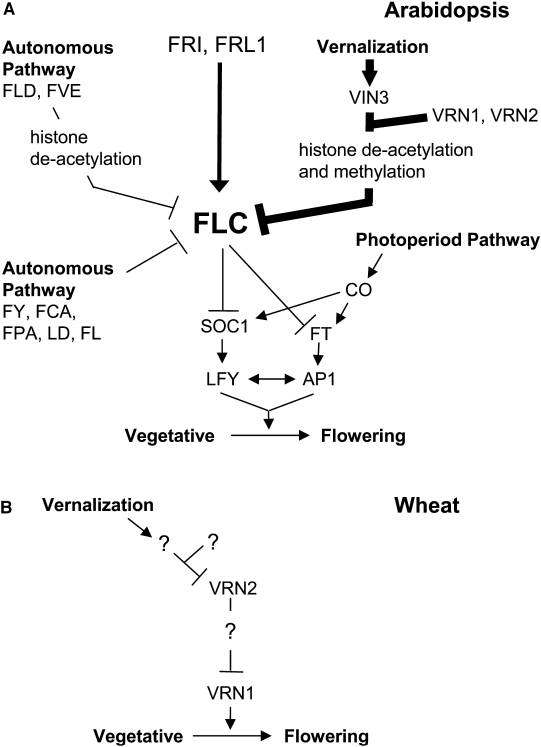


Figure 3 – The regulatory pathways involved in triggering flowering in *Arabidopsis thaliana* (Amasino 2004). FRL2 (not shown) is associated with promotion of *FLC* expression (Finnegan et al. 2005) and TSF (not shown) which is regulated in a similar fashion to *FT* (Hiraoka et al. 2013). Additional genes critical to the process relating to the PHD-PRC2 complex during vernalisation have also not been shown.

*FT* (*TSF*), and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*), and increased FT, TSF and SOC1 levels in turn suppresses *LEAFY* (*LFY*) and *APETALA1* (*AP1*) expression, two primary promoters of floral apical meristem growth (see Fig. 3; Amasino 2004). Upon *FLC* expression induction, the VERNALISATION2(*VRN2*)/Plant Homeodomain Polycomb Repression Complex22 (PHD-PRC2) complex that consists of *VRN2* and PHD-PRC2 proteins, CURLY LEAF (CLF),SWINGER(SWN), and FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) (Köhler & Villar 2008), is constitutively bound to the *FLC* locus. The binding of the VRN2/PHD-PRC2 complex to *FLC* maintains the locus in an open confirmation, allowing for transcription machinery access to *FLC and FLC expression*, via histone H3 acetylation (of what??) During vernalisation, *VERNALISATION INSENSITIVE3* (*VIN3*) levels increase. VIN3, along with VERNALIZATION5/VIN3-LIKE1 (VEL1) and VRN5, bind to the PHD-PCR2 complex to promote histone H3 deacetylation, and VRN1 and VRN2-directed methylation of H3K9 and H3K27 (Fig. 3B). Histone methylation of the FLC locus closes the open confirmation of FLC, blocking transcription machinery access to FLC, to repress FLC expression. This epigenetic repression of FLC is irreversible, and ensures that the vernalised plant transitions from vegetative to reproductive growth (Levy et al. 2002; Sung & Amasino 2004). Reduced *FLC* levels increases *SOC1* and *FT* levels, and SCO1 and FT in turn promote *LFY* and *AP1* expression. The External Coincidence Model (also known as the Photoperiod Pathway) characterises transition to flowering without necessarily

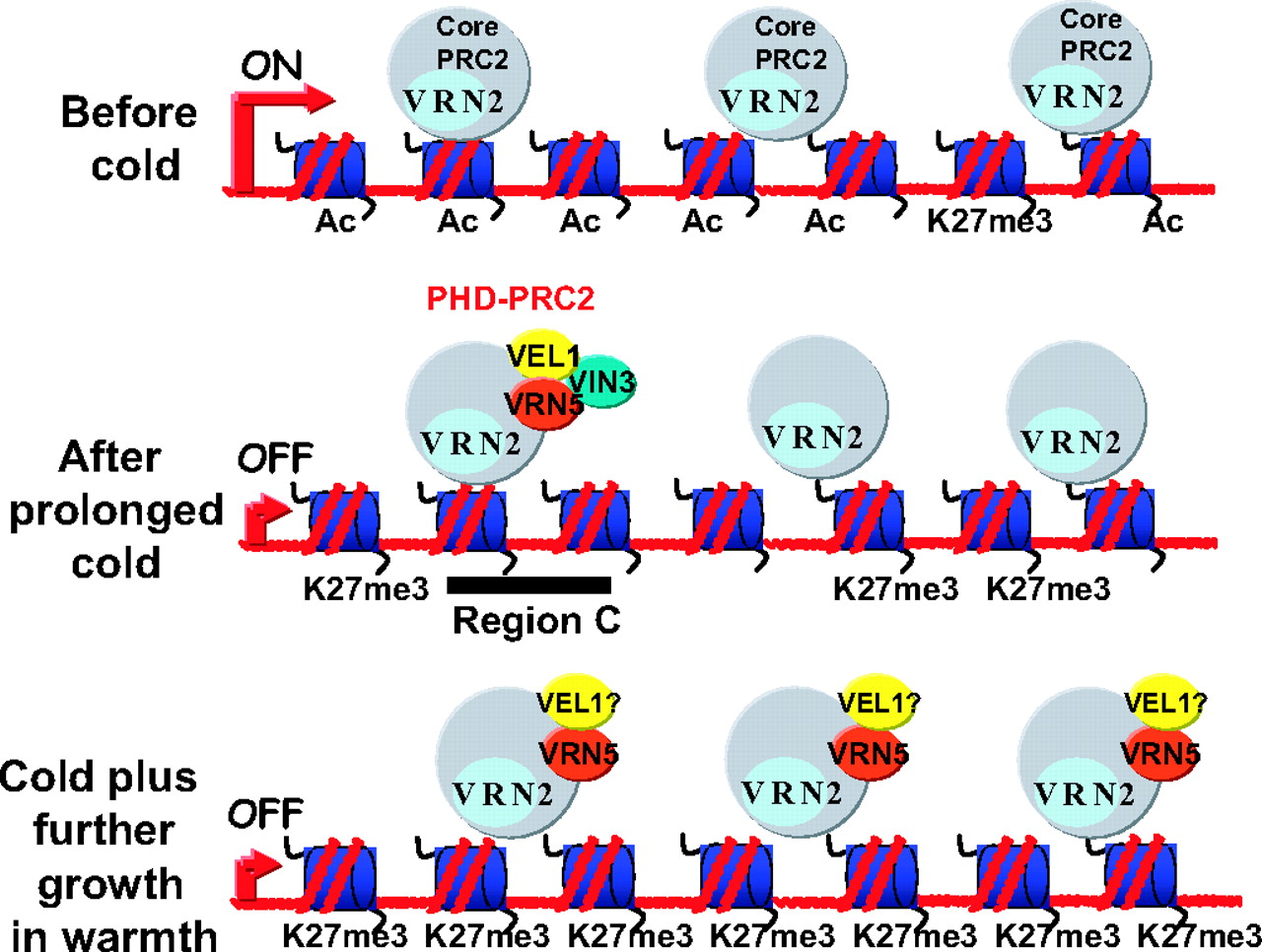


Figure 4 – PHD-PCR2 and VRN5 complex and its interaction before, during and after vernalisation (De Lucia et al. 2008)

The External Coincidence Model, also referred to as the Photoperiod Pathway, mediates the transition to flowering by exposure to increasing day length, and without the requirement of a period of vernalisation (Hayama & Coupland 2004). Increased periods of day light promotes the expression of *CONSTANS* (*CO*), and CO in turn overrides the repression effects of FLC, via CO-mediated activation of of *FT* and *SOC1* expression (Golembeski et al. 2014). Even without exposure to cold or increasing day length, the Autonomous Pathway can trigger the floral transition of *A. thaliana* by down regulating *FLC* expression. *FCA* with *FY*, *FLA* and *FPK* are all independently involved with *FLC* RNA processing, *FLD* and *FVE* deacetylate histones at the *FLC* locus. The result is similar gene regulation as seen with vernalisation but initiated and progressing at a much slower rate (Simpson 2004).

In *A. thaliana*, summer ecotypes possess allelic variations, but not necessarily in specifically vernalisation genes. The *A. thaliana* ecotype Landsberg *erecta* (Ler-0) contains an allele of *FLC* which is unresponsive to up regulation by *FRI*, meaning there is no repression of *FT*, leading to early flowering of L*er*-0. Therefore, the necessity to decrease *FLC* expression by vernalisation and increased expression of

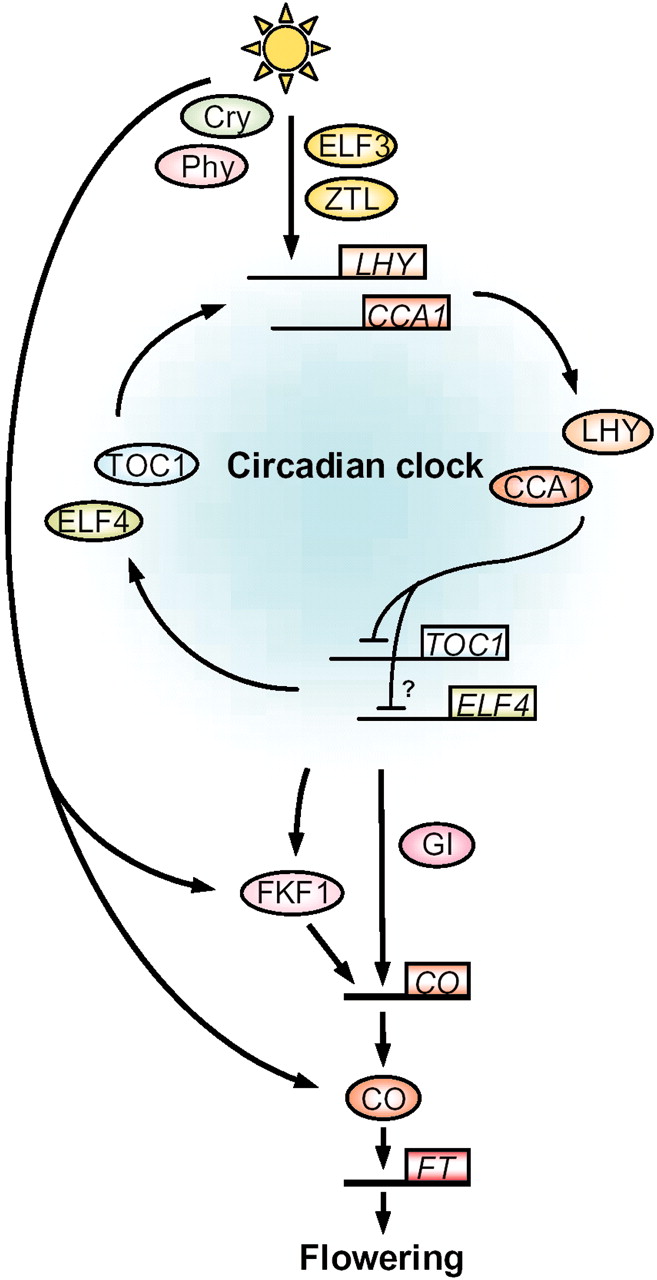


Figure 5 – The External Coincidence Model in *A. thaliana*. This model bypasses the down regulating effect of *FLC* on *FT*, increasing *FT* expression and allowing the transition to flowering. (Hayama & Coupland 2004)

*VRN1* and *VRN2* (or the necessity for long day conditions to increase the expression of *LD*) to down regulate *FLC* is no longer required (Michaels & Amasino 1999).There are many other *A. thaliana* ecotypes that can be categorised as “summer” and “winter” based on their response to vernalisation (Nordborg & Bergelson 1999).

## From here needs a lot of additional work by you before I can comment – please amend.

## Vernalisation in Cereals

In cereals (e.g. wheat, barley) and many other monocots (e.g. *Brachypodium distachyon*), the vernalisation pathway is not as comprehensively understood, but differs greatly from *A. thaliana*. While there are many conserved mechanisms between cereals and *A. thaliana* (such as the Photoperiod Pathway) (Griffiths et al. 2003), the most notable difference in cereals is the absence of a known homolog for *AtFLC* (see figure 6 below). Instead, *VRN1* (a MADS-box transcription factor) and *VRN2* (a zinc-finger motif with CCT domain, which in cereals is distinct from *AtVRN2*), along with *FT*, are responsible for regulation of flowering in cereals (Yan et al. 2004). Before wintering, *VRN2* is expressed at higher levels, repressing expression of *FT* and maintaining vegetative growth (Ream et al. 2014). In “winter” barley (*Hordeum vulgare*) varieties, *HvVRN1* expression is up-regulated by vernalisation, where *HvVRN2* expression, along with *CO,* is induced primarily by increasing day length, resulting in increased *LD* expression (Trevaskis et al. 2006). As the cereal is vernalised, similar to *FLC* in *A. thaliana*, histone demethylation and acetylation take place in *VRN1*, resulting in the increase in *VRN1* expression. Once temperatures and day length increase with the onset of spring, the stable epigenetic changes to *VRN1* remain, continuing to represses expression of *VRN2*. Combined with increasing day length and the subsequent repression of *VRN2*, expression of *FT* increases and the cereal transitions to flowering (see figure 7 below). Many spring cereals contain alleles of *VRN2* such thatit does not suppress *FT*, therefore the requirement for vernalisation is no longer present. In these varieties, *FT* expression is never suppressed, so when day length conditions are correct, this triggers expression of *CO* which in turn triggers *FT* expression and allows transition to flowering.

It is interesting to note that the response to vernalisation is in rice (*Oryza sativa*) differs substantially due to its acclimatisation to tropical environments. Exposure of rice to cold conditions during development (significantly warmer and for shorter periods than seen in vernalisation) leads to sterility and damage to developing flower and seed heads, resulting in significant crop losses (Oliver et al. 2007).

* Nature Communications – ODDSOC 2, Cereal synteny
* NB: in Trevaskis 2006, it talks about zinc finger proteins *HvCCTa* and *HvCCTb* with regard to vernalisation. Any comparison to the PHD-PRC2 complex in AraTha?

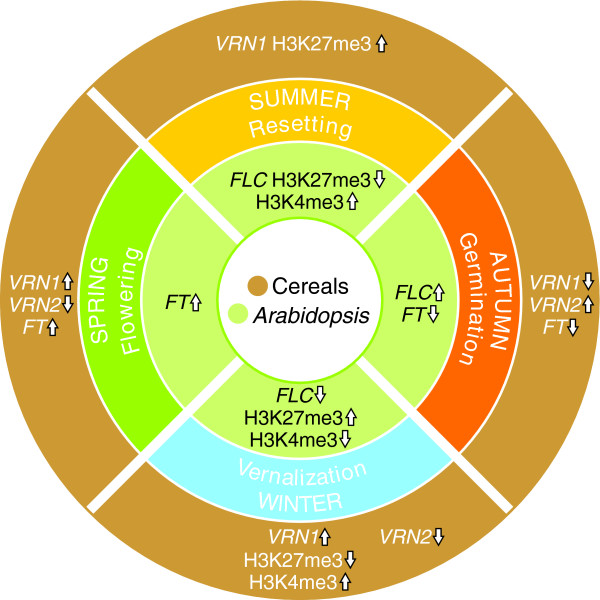


Figure 6 – Comparison of gene expression and histone modifications between vernalisation-sensitive *Arabidopsis thaliana* and “winter” cereals, the most notable difference is the absence of FLC in *A. thaliana* (Dennis & Peacock 2009).

***PPD1***

***CO***

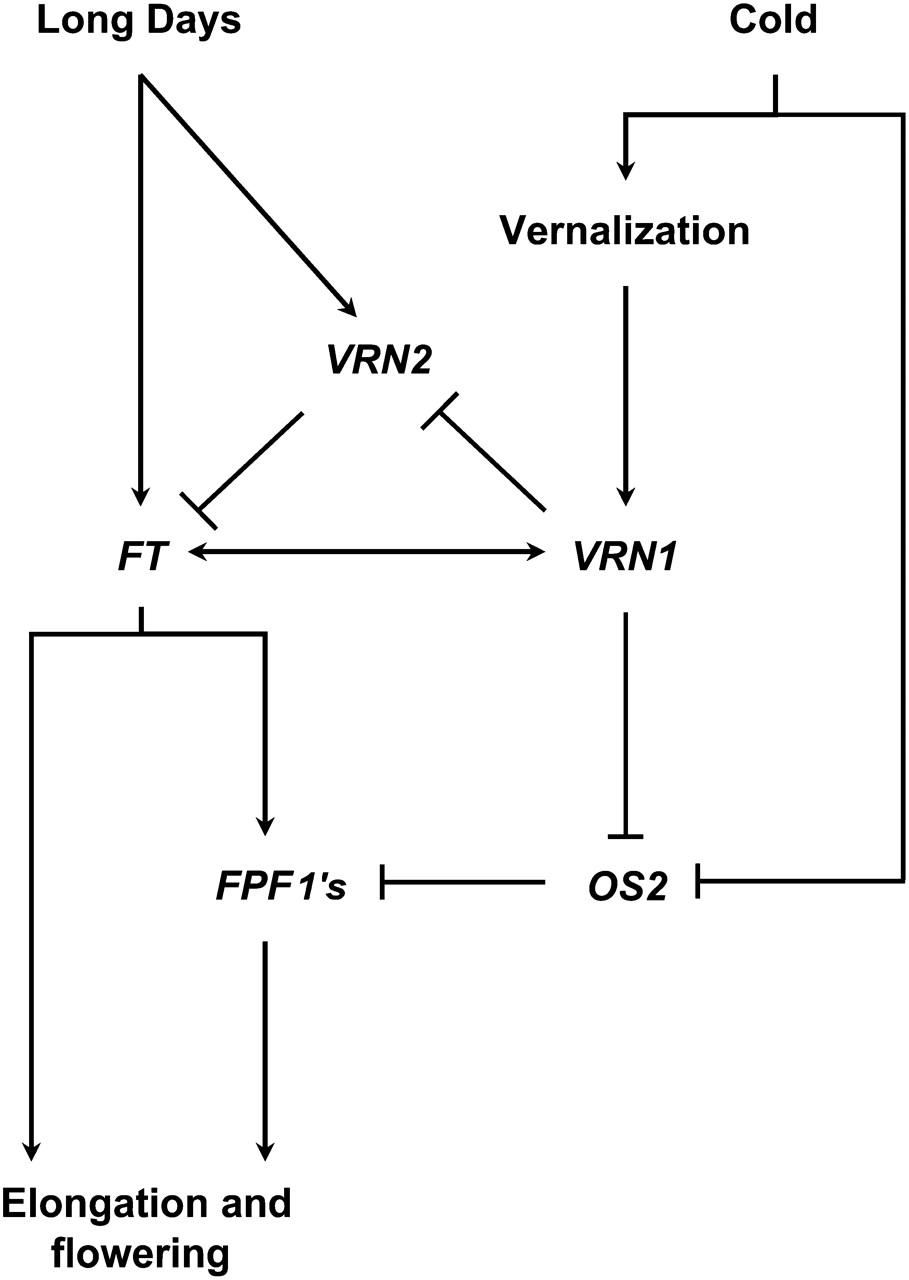


Figure 7 – Vernalisation gene regulatory pathway involved in triggering flowering in winter cereals. While the vernalisation response is different, response to day length cues contains similarities with *A. thaliana* (Modified from (Greenup et al. 2010) and (Trevaskis et al. 2007))

## Vernalisation in Beets

Sugar Beet (*Beta vulgaris*) has also evolved an alternate method of regulating flowering time based on day length and wintering. In a study of wild “Sea Beet” (*B vulgaris* ssp. *maritimavia*) distributed around the coast of France, it was found that the necessity for vernalisation increases the further north they were found (Boudry et al. 2002). However, in more recent studies (Pin et al. 2012), it has been found that vernalisation mechanisms in beet are distinct from *A. Thaliana* and cereals (see figure 8 below). Rather than a single *FT* gene that triggers flowering time, as seen in *A. thaliana*, two *FT* genes While *B. vulgaris* contains a number of genes consistent with the Photoperiod Pathway in *A. thaliana* , the critical gene for control of transition to flowering is *BOLTING TIME CONTOL 1* (*BvBTC1*), up-regulated by *LD* expression (Pin et al. 2012).

BTC1 controls the interaction between BvFT1 and BvFT2, leading to control of flowering time

* This might be a good spot to speak about FT-like and other “x-likes” and homologues of genes in both similar and differing species.

## Determining Factors Affecting Vernalisation

There are a number of differing and redundant pathways in *A. thaliana*, cereals and *B. Vulgaris* that can trigger flowering, such that a mutation in the vernalisation pathway may not necessarily be the only way that the transition to flowering may be interfered with. For example, a mutation that removes *CO* regulators may cause the plant to lose the requirement for lengthening light as a flowering trigger and result in a fast transition to flowering, regardless of the plants existing requirement for vernalisation. Recent *A. thaliana* research (Nakamura et al. 2014) shows that higher levels of a diacyl glyceride (DAG) comprising of phosphatidylcholine (PC) and oleic acid (18:1) in the shoot apical meristem accelerates flowering, and vice versa. This research also shows other DAGs containing PC binding to *FT*. This indicates that levels of certain specific lipids in plant tissues could also be another contributing factor of plant flowering. However, at this point, it is unclear if and how lipid levels in plant tissues affect a plant’s response to vernalisation.

While safflower, *A. thaliana* and *B. vulgaris* are all dicots, there are enough differences in vernalisation pathways between *A. thaliana* and *B. Vulgaris* that it is plausible there are a number of factors in safflower, possibly even across the

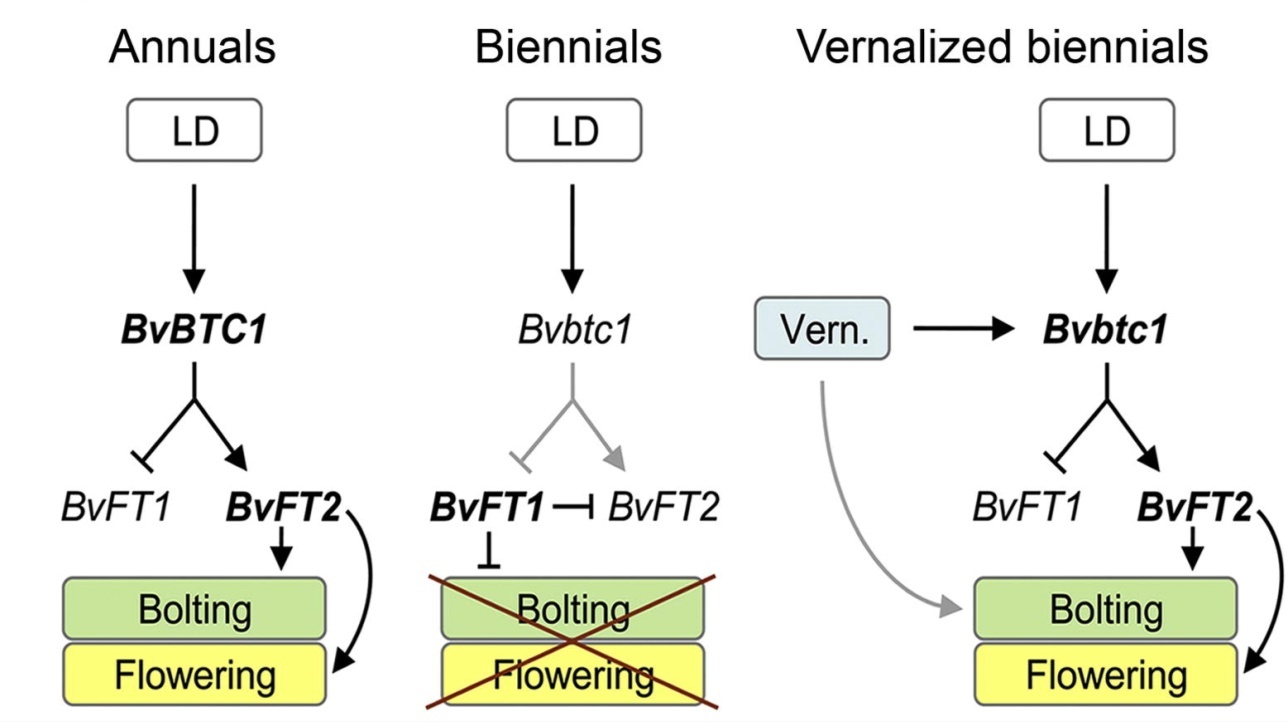


Figure 8 – Regulatory mechanisms for flowering and vernalisation in *Beta vulgaris* (Pin et al. 2012).

Asteraceae as a family, that set them apart from other dicots with regards to vernalisation. If this is the case, these confounding factors will impact on the elucidation of which genes are critical in the vernalisation pathway, especially when comparing a vernalisation sensitive and vernalisation resistant safflower cultivar.

* Your sequencing sequence requires a lot of cleaning up – its very draft-ish currently – please keep working on this section

# Sequencing Methodologies

The availability of genomic and transcriptomic resources has advanced biological research in leaps and bounds. Since the 1970s, genetic resources have allowed elucidation of genes and transcriptional factors involved in not only vernalisation, but almost every major metabolic pathway in plants, animals and microbes. Advancements in sequencing technology have increased the quantity and quality of sequencing data that can be obtained from a sample by orders of magnitude in a fraction of the time.

## Sanger Sequencing

One of the first sequencing techniques was the “chain termination” method originally developed by (Sanger & Coulson 1975) and colloquially named “Sanger Sequencing”. It involves cloning a sequence of interest (genomic or transcriptomic) in four separate Polymerase Chain Reactions (PCRs), each containing a single type of di-deoxynucleotide triphosphate (ddNTP). During elongation, if the polymerase binds a ddNTP to the nucleotide chain instead of a deoxynucleotide triphosphate (dNTP), the absence of a 3'-hydroxyl group on the ddNTP prevents another dNTP or ddNTP from bonding. When electrophoresed over polyacrylamide gel and fluoresced, the result is a number of varying length bands in each column corresponding to the order of nucleotides. Many significant genomes have been constructed using Sanger Sequencing, including *Caenorhabditis elegans* (Wilson, 1999), *A. thaliana* and the human genome. The disadvantage with Sanger sequencing is the time and cost involved. The human genome (approx 3.4 Gbp, diploid) cost approximately USD $13 billion dollars and took hundreds of labs across the world thirteen years to complete, at a cost of around $4 per nucleotide (International Human Genome Sequencing Consortium 2004). Today, data generated by a single lane of an Illumina HiSeq 2500 sequencer calculates to approximately $0.25 per million nucleotides. However, per nucleotide, while Sanger Sequencing is the oldest, most time consuming and expensive method of sequencing, it is still considered the highest quality of sequencing currently available.

## Expressed Sequence Tags (ESTs)

Expressed Sequence Tags are short cDNA fragments between 200 bp and 800 bp in length representing an mRNA transcript. They are created by randomly cloning and sequencing transcripts in a sample using PCR. The result is an “expression profile” (i.e. a unique set of genes expressed in a tissue type or environmental condition) of which each sequence within the set uniquely identifies a gene expressed. These expression profiles can then be compared against one another to identify differentially expressed genes, such as between diseased and healthy tissue (Rezvani et al. 2000). One limitation with ESTs is the need for *a priori* knowledge of transcripts to characterise genes. A novel EST is difficult to identify if there is no sequence homology against a known transcript, meaning a transcript will be known to be expressed but without necessarily having any knowledge of its function. Another limitation is that identification of rare transcripts, such as those found in the human brain, can be difficult without the use of unique primer sets, which require either *a priori* knowledge of the transcript (Adams et al. 1991) or require additional molecular techniques, such as primer degredation. If a rare transcript and an abundant transcript both contain the same PCR primers, the chance of primers binding to the rare transcript over the abundant one will be low, meaning difficulty separating the rare transcript due to the high signal produced by the abundant transcript.

While newer Next Generation Sequencing (NGS) technologies surpass ESTs with regard to the quantity and quality of transcript expression, because ESTs are directly cloned from extracted mRNA and do not require assembly of fragments via algorithm, they often complement *de novo* assemblies as a way verifying the quality and accuracy of an assembly.

## Microarrays

Microarrays are constructed by binding millions of oligonucleotide fragments bound to a glass slide either by photochemistry (Fodor et al. 1993) or technology similar to inkjet printer (Allain et al. 2001). Microarrays can consist of RNA fragments representing key genes, tissues or sets of genes for medical diagnostics, genomic locations to observe nucleotide polymorphisms, methylation arrays (CHiP-on-chip) to examine gene regulation and methylation as well sets of tiling arrays for sequencing whole genomes. Each microarray can only contain 6.5 million olignucleotides on its surface, which restricts the application of microarrays to specific medical diagnostics, use in characterising small genomes such as bacteria, smaller transcriptomes or sets of common transcripts. Sets of microarrays can also be used for sequencing smaller eukaryotic genomes such as *A. thaliana*. Similar to ESTs, the requirement for *a priori* knowledge of oligonucleotides for microarray construction and analysis decreases the capability for use in discovering *de novo* or low count transcripts, though similar related species can be used as an adjunct.

## Next Generation Sequencing (NGS)

In the last 15 years, Next Generation Sequencing(NGS) has expanded the quantity of information available to scientists by leaps and bounds. Depending on the type of sequencing required (e.g. DNA, mRNA, sRNA, microRNA), while the library preparation steps change, the process of sequencing comprises of randomly digesting and filtering samples by size. Illumina are one of the most common sequencing

* Illumina vs pyrosequencing vs newer tech (ion torrent, Moleculo)

Single/paired end, mate pair and scaffolding

NGS processing – De-brujin graphs vs (kanga methodology?)

## Assessing the Quality of a *de novo* Assembly

Regardless of whether a *de novo* assembly is genomic or transcriptomic, because there is no reference of which to make a comparison, assessing the quality of a *de novo* assembly is challenging. A number of techniques have been developed to address this problem, with varying success. Calculating the n50 of an assembly (i.e. the length of the contig containing 50% of the nucleotides of the entire sorted assembly) gives an initial indication of the quality of an assembly, it does not provide insight on any misassembly or chimeric contigs produced by the asssembly algorithm. <n50 and skewed metrics example> The most effective way for assessing transcriptome is aligning good quality *known* sequences (such as transcripts that have been sequenced by Sanger techniques) or ESTs against the *de novo* assembly. The longer and more numerous the alignments, the better the quality of the assembly. Where reference sequences do not exist, the Core Eukaryotic Genes Mapping Approach (CEGMA) can be used instead. By aligning these 248 most common eukaryotic transcripts and aligning them against the *de novo* genome, this can indicate assembly quality as the higher number and score of alignments of *de novo* contigs against conserved genes, the better the quality of the assembly (Parra et al. 2007).

# Aims of this research project:

Set your Aims out in parts, so that your examiner can be confident that each aim will allow for the construction of a results chapter in your Thesis.

Also be a lot more positive here and talk up this project – it is very exciting research and this needs to be conveyed to your examiners!

The primary aim of this research project is to investigate how vernalisation results in phenotypic variation in safflower through examination of vernalisation sensitive and insensitive cultivars*.* There are a number of approaches to achieve this outcome. The first is the generation and analysis of *in silico* resources, namely a draft *de novo* genome and transcriptome (in combination with data from existing published transcriptomes), including differential expression profiles within different cultivars, plant phenotypes and tissues. *In silico* putative transcripts differentially expressed in vernalised tissues will be analysed to determine their candidature as vernalisation genes, as well as genes originating from other plant species sharing sequence similarity to *de novo* transcripts. *In silico* candidate genes will be verified or refuted using molecular biology techniques. The results of this research will then influence traditional breeding techniques as well as genetic manipulation of safflower cultivars with a goal to diversify the regions safflower can be grown in Australia.

# References

Adams, M.D. et al., 1991. Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science*, 252(5013), pp.1651–1656.

Allain, L. et al., 2001. Microarray sampling-platform fabrication using bubble-jet technology for a biochip system. *Fresenius’ Journal of Analytical Chemistry*, 371(2), pp.146–150. Available at: http://dx.doi.org/10.1007/s002160100962.

Amasino, R., 2004. Vernalization, Competence, and the Epigenetic Memory of Winter. *The Plant Cell Online*, 16(10), pp.2553–2559. Available at: http://www.plantcell.org/content/16/10/2553.short.

Boss, P.K. et al., 2004. Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting. *The Plant Cell Online*, 16(suppl 1), pp.S18–S31. Available at: http://www.plantcell.org/content/16/suppl\_1/S18.short.

Boudry, P., McCombie, H. & Dijk, H. van, 2002. Vernalization Requirement of Wild Beet Beta vulgaris ssp. maritima: Among Population Variation and Its Adaptive Significance. *Journal of Ecology*, 90(4), pp.693–703. Available at: http://www.jstor.org/stable/3072271.

Chempro Technoatvation Pvt. Ltd., Fatty Acid Composition of some Major Oils. Available at: http://www.chempro.in/fattyacid.htm [Accessed August 5, 2014].

Chouard, P., 1960. Vernalization and its Relations to Dormancy. *Annual Review of Plant Physiology*, 11(1), pp.191–238. Available at: http://dx.doi.org/10.1146/annurev.pp.11.060160.001203.

Dennis, E. & Peacock, W.J., 2009. Vernalization in cereals. *Journal of Biology*, 8(6), p.57. Available at: http://jbiol.com/content/8/6/57.

Finnegan, E.J. et al., 2005. The downregulation of FLOWERING LOCUS C (FLC) expression in plants with low levels of DNA methylation and by vernalization occurs by distinct mechanisms. *The Plant Journal*, 44(3), pp.420–432. Available at: http://dx.doi.org/10.1111/j.1365-313X.2005.02541.x.

Fodor, S.P.A. et al., 1993. Multiplexed biochemical assays with biological chips. *Nature*, 364, pp.555–556.

Garnatje, T. et al., 2006. Genome Size Variation in the Genus Carthamus (Asteraceae, Cardueae): Systematic Implications and Additive Changes During Allopolyploidization. *Annals of Botany*, 97(3), pp.461–467. Available at: http://aob.oxfordjournals.org/content/97/3/461.abstract.

Gecgel, U. et al., 2007. Fatty Acid Composition of the Oil from Developing Seeds of Different Varieties of Safflower (Carthamus tinctorius L.). *Journal of the American Oil Chemists’ Society*, 84(1), pp.47–54. Available at: http://dx.doi.org/10.1007/s11746-006-1007-3.

Golembeski, G.S. et al., 2014. Chapter One - Photoperiodic Flowering Regulation in Arabidopsis thaliana. In F. F. B. T.-A. in B. Research, ed. *The Molecular Genetics of Floral Transition and Flower Development*. Academic Press, pp. 1–28. Available at: http://www.sciencedirect.com/science/article/pii/B9780124171626000018.

Greenup, A.G. et al., 2010. ODDSOC2 Is a MADS Box Floral Repressor That Is Down-Regulated by Vernalization in Temperate Cereals. *Plant Physiology*, 153(3), pp.1062–1073. Available at: http://www.plantphysiol.org/content/153/3/1062.abstract.

Griffiths, S. et al., 2003. The Evolution of CONSTANS-Like Gene Families in Barley, Rice, and Arabidopsis. *Plant Physiology*, 131(4), pp.1855–1867. Available at: http://www.plantphysiol.org/content/131/4/1855.abstract.

Hayama, R. & Coupland, G., 2004. The Molecular Basis of Diversity in the Photoperiodic Flowering Responses of Arabidopsis and Rice. *Plant Physiology*, 135(2), pp.677–684. Available at: http://www.plantphysiol.org/content/135/2/677.short.

Hiraoka, K. et al., 2013. The Florigen Genes FT and TSF Modulate Lateral Shoot Outgrowth in Arabidopsis thaliana. *Plant and Cell Physiology*, 54(3), pp.352–368. Available at: http://pcp.oxfordjournals.org/content/54/3/352.abstract.

Ingram, D.S., Vince-Prue, D. & Gregory, P.J., 2008. *Science and the Garden: The Scientific Basis of Horticultural Practice*, Wiley-Blackwell, West Sussex.

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), pp.931–945. Available at: http://dx.doi.org/10.1038/nature03001.

Işigigür, A., Karaosmanoglu, F. & Aksoy, H.A., 1995. Characteristics of safflower seed oils of turkish origin. *Journal of the American Oil Chemists’ Society*, 72(10), pp.1223–1225. Available at: http://dx.doi.org/10.1007/BF02540994.

Johnson, R.C., Dajue, L. & Bradley, V., 2006. Autumn growth and its relationship to winter survival in diverse safflower germplasm. *Canadian Journal of Plant Science*, 86(3), pp.701–709. Available at: http://dx.doi.org/10.4141/P05-104.

Knights, S., 2010. *Raising the bar with better safflower agronomy*, Available at: http://www.australianoilseeds.com/\_\_data/assets/pdf\_file/0010/6949/GDRC\_Raising\_the\_Bar\_-\_Better\_Safflower\_Agronomy.pdf.

Knowles, P.F., 2012. *Safflower in California: The Paulden F. Knowles personal history of plant exploration and research on evolution, genetics, and breeding* P. E. McGuire, A. B. Damania, & C. O. Qualset, eds., Department of Plant Sciences, University of California, Davis.

Köhler, C. & Villar, C.B.R., 2008. Programming of gene expression by Polycomb group proteins. *Trends in Cell Biology*, 18(5), pp.236–243. Available at: http://www.sciencedirect.com/science/article/pii/S0962892408000809.

Levy, Y.Y. et al., 2002. Multiple Roles of Arabidopsis VRN1 in Vernalization and Flowering Time Control. *Science*, 297(5579), pp.243–246. Available at: http://www.sciencemag.org/content/297/5579/243.abstract.

Li, H. et al., 2012. De Novo Transcriptome of Safflower and the Identification of Putative Genes for Oleosin and the Biosynthesis of Flavonoids. *PLoS ONE*, 7(2), p.e30987. Available at: http://dx.doi.org/10.1371/journal.pone.0030987.

Li, H. et al., 2011. Investigation of the microRNAs in safflower seed, leaf, and petal by high-throughput sequencing. *Planta*, 233(3), pp.611–619. Available at: http://dx.doi.org/10.1007/s00425-010-1327-2.

De Lucia, F. et al., 2008. A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proceedings of the National Academy of Sciences*, 105(44), pp.16831–16836. Available at: http://www.pnas.org/content/105/44/16831.abstract.

Lulin, H. et al., 2012. The First Illumina-Based De Novo Transcriptome Sequencing and Analysis of Safflower Flowers. *PLoS ONE*, 7(6), p.e38653. Available at: http://dx.doi.org/10.1371/journal.pone.0038653.

Macindoe, S.L. & Brown, C., 1968. *Wheat breeding and varieties in Australia* C. Walken, ed., Sydney.

Michaels, S.D. & Amasino, R.M., 1999. FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *The Plant Cell Online*, 11(5), pp.949–956. Available at: http://www.plantcell.org/content/11/5/949.abstract.

Nakamura, Y. et al., 2014. Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nat Commun*, 5. Available at: http://dx.doi.org/10.1038/ncomms4553.

Nordborg, M. & Bergelson, J., 1999. The Effect of Seed and Rosette Cold Treatment on Germination and Flowering Time in Some Arabidopsis thaliana (Brassicaceae) Ecotypes. *American Journal of Botany*, 86(4), pp.470–475. Available at: http://www.jstor.org/stable/2656807.

Oliver, S.N., Dennis, E.S. & Dolferus, R., 2007. ABA Regulates Apoplastic Sugar Transport and is a Potential Signal for Cold-Induced Pollen Sterility in Rice. *Plant and Cell Physiology*, 48(9), pp.1319–1330. Available at: http://pcp.oxfordjournals.org/content/48/9/1319.abstract.

Parra, G., Bradnam, K. & Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9), pp.1061–1067. Available at: http://bioinformatics.oxfordjournals.org/content/23/9/1061.abstract.

Pin, P.A. et al., 2010. An Antagonistic Pair of FT Homologs Mediates the Control of Flowering Time in Sugar Beet. *Science*, 330(6009), pp.1397–1400. Available at: http://www.sciencemag.org/content/330/6009/1397.abstract.

Pin, P.A. et al., 2012. The Role of a Pseudo-Response Regulator Gene in Life Cycle Adaptation and Domestication of Beet. *Current Biology*, 22(12), pp.1095–1101. Available at: http://www.cell.com/current-biology/abstract/S0960-9822(12)00394-6.

Ream, T.S. et al., 2014. Interaction of Photoperiod and Vernalization Determines Flowering Time of Brachypodium distachyon. *Plant Physiology*, 164(2), pp.694–709. Available at: http://www.plantphysiol.org/content/164/2/694.abstract.

Rezvani, M. et al., 2000. Apoptosis-related genes expressed in cardiovascular development and disease: an EST approach. *Cardiovascular Research*, 45(3), pp.621–629.

Salisbury, F.B. & Ross, C.W., 1992. *Plant Physiology*, Wadsworth Publishing Company. Available at: http://books.google.com.au/books?id=F6KQGAAACAAJ.

Sanger, F. & Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), pp.441–448. Available at: http://www.sciencedirect.com/science/article/pii/0022283675902132.

Sheldon, C.C. et al., 2000. The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). *Proceedings of the National Academy of Sciences*, 97(7), pp.3753–3758. Available at: http://www.pnas.org/content/97/7/3753.abstract.

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*, 7(5), pp.570–574. Available at: http://www.sciencedirect.com/science/article/pii/S1369526604000949.

Sung, S. & Amasino, R.M., 2004. Vernalization and epigenetics: how plants remember winter. *Current Opinion in Plant Biology*, 7(1), pp.4–10. Available at: http://www.sciencedirect.com/science/article/pii/S1369526603001419.

Trevaskis, B. et al., 2006. HvVRN2 Responds to Daylength, whereas HvVRN1 Is Regulated by Vernalization and Developmental Status. *Plant Physiology*, 140(4), pp.1397–1405. Available at: http://www.plantphysiol.org/content/140/4/1397.abstract.

Trevaskis, B. et al., 2007. The molecular basis of vernalization-induced flowering in cereals. *Trends in Plant Science*, 12(8), pp.352–357. Available at: http://www.sciencedirect.com/science/article/pii/S1360138507001562.

United Nations, 2014. FAOSTAT. Available at: http://data.fao.org/dataset-data-filter?entryId=29920434-c74e-4ea2-beed-01b832e60609&tab=data.

Wellensiek, S.J., 1962. Dividing Cells as the Locus for Vernalization. *Nature*, 195(4838), pp.307–308. Available at: http://dx.doi.org/10.1038/195307a0.

Yan, L. et al., 2004. The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science*, 303(5664), pp.1640–1644. Available at: http://www.sciencemag.org/content/303/5664/1640.abstract.

Yazdi-Samadi, B. & Zali, A.A., 1979. Comparison of Winter- and Spring-Type Safflower1. , pp.783–785. Available at: https://www.crops.org/publications/cs/abstracts/19/6/783.

Zohary, D. & Hopf, M., 1993. *Domestication of Plants in the Old World* 2nd Ed., Oxford Scientific Publications.