Lit Review

Table of Contents:

Abbreviations 3

Gene Abbreviations 5

Safflower – History and Characteristics 7

Genetic and molecular characteristics of safflower 7

Vernalisation Response in Safflower 8

Vernalisation 9

History of Vernalisation 10

The Biology of Vernalisation 10

The *Arabidopsis thaliana* vernalisation pathway 12

Vernalisation pathways in cereals 14

Vernalisation in Sugar Beet 16

Vernalisation in other *Asteraceae* 17

Determining Factors Affecting Vernalisation 17

Mapping of Pathways and Characterising Genes 19

Traditional Crossing Methods 19

Characterisation using Forward Genetics 19

Characterisation using Reverse Genetics 20

Sequencing Methodologies 20

Sanger Sequencing 20

Expressed Sequence Tags (ESTs) 21

Microarrays 22

Next Generation Sequencing 22

Creating a *de novo* assembly 23

Assessing the Quality of a *de novo* Assembly 25

Aims of this research project: 25

References 27

# 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)

*Arabidopsis – Arabidopsis thaliana* (thale cress)

*B. vulgaris – Beta vulgaris* (sugar beet)

bp – base pair

*C. elegans – Caenorhabditis elegans*

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

DAG – diacyl glyceride

ddNTP – di-deoxynucleotide triphosphate

dicot – dicotyledonous species

dNTP – deoxynucleotide triphosphate

EST – Expressed Sequence Tag

F# – the generation of progeny (#) after a population has been created by the crossing of two plant varieties

Gbp – gigabase pair (1 Gbp = 1,000,000,000 bp)

Gnt – giganucleotide (1 Gnt = 1,000,000,000 nt)

*H. vulgare – Hordeum vulgare* (barley)

HDAC – Histone deacetylase complex

kbp – kilobase pair (1 kbp = 1,000 bp)

knt – kilonucleotide (1 knt = 1,000 nt)

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

M# – the generation of mutated progeny (#) after genomic mutations have been introduced to a population

Mbp – megabase pair (1 Mbp = 1,000,000 bp)

Mnt – meganucleotide (1 Mnt = 1,000,000 nt)

Monocot – monocotyledous species

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 – tonne (metric)

*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

*FLA – fasciclin-like arabinogalactan*

*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*

*VEL –* VERNALISATION 5/VIN3-LIKE

*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* – An abbreviation in *CAPITAL ITALICS* represents a gene or mRNA. An abbreviation in CAPITALS represents a translated protein i.e. “*FLC*”is the gene or mRNA, “FLC” is the protein

*Gene Abbreviations Note 2* – 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. *Arabidopsis* (*At*), *Beta vulgaris* (*Bv*)

*Gene Abbreviations Note 3* – When a gene has a recessive allele, it will be indicated as lower case i.e. *FLOWERING TIME – FT* (dominant, normal) verses. *ft* (recessive)

* Nomenclature for plant genes, mRNAs and proteins
* Genes are capital italics (e.g., *FLC*)
* mRNAs are in capital italics (e.g., *FLC*)
* proteins are in capitals (e.g., FLC)

# Safflower – History and Characteristics

Safflower (*Carthamus tinctorius* L.) belongs to the *Asteraceae* family of flowering plants and is native to the regions of the eastern and southern Mediterranean, the Middle East and India (REF). It has been cultivated in these regions for thousands of years, with safflower arrangements and safflower based dyes found in many archaeological sites in Egypt and Mesopotamia, including Pharaoh Tutankhamen’s tomb (Zohary & Hopf 1993). Safflower was originally cultivated for its edible seeds and dyes extracted from its vibrant yellow and orange flowers. Safflower seed oil is comparable to sunflower seed oil (*Helianthus annuus* L.), both have similar melting points and fatty acid composition (Chempro Technovation Pvt. Ltd. n.d.). However, safflower oil is less susceptible to oxidation. For this reason, it is often used as a base for varnishes and oil based paints and as an effective industrial lubricant. After the oil has been extract from safflower seeds, the residual meal can be used as animal feed (Işigigür et al. 1995; Gecgel et al. 2007). In 2012, the global harvest of safflower seed was 833,793 tonnes (t) (see figure 1 below), with 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 primarily grown as a “break crop”, to break up hard clay pans or to remove excess water from soils before the cultivation of traditional crops, such as wheat and barley (Knights 2010).

## Genetic and molecular characteristics of safflower

Safflower is a dicotyledonous plant (dicot) with a diploid genome of 12 chromosomal pairs. The approximate haploid genome size (i.e. the estimated size of 12 unpaired chromosomes) in four safflower cultivars, namely “Ljubljana”, “Uzbekistan”, “S-2190” and “Huesca” has been calculated as 1.34 Gbp (1 Gbp = one Gigabase pair = 1,000,000,000 base pairs), 1.38 Gbp, 1.39 Gbp and 1.40 Gbp respectively (Garnatje et al. 2006). However, to date, despite there being publicly accessible 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 for Biotechnology Information (NCBI), currently, there is no publicly accessible genome for safflower, or indeed any other *Asteraceae* species, exists, despite the characterisation of major metabolic pathways in specific *Asteraceae*. For example, in safflower, research into the fatty acid synthesis pathway has allowed the characterisation of eleven members of the fatty acid desaturase gene family, providing insight on the metabolic synthesis of fatty acid in safflower oil (Cao et al. 2013). Another example is research to characterise the *LOSS OF APOMEOSIS* (LOA) and *LOSS OF PARTHENOGENESIS* (LOP) loci in Hieracium is unlocking the genetic mechanisms and regulatory pathways that underpin apomictic behaviour (Koltunow et al. 2011). While research on these molecular pathways has been quite

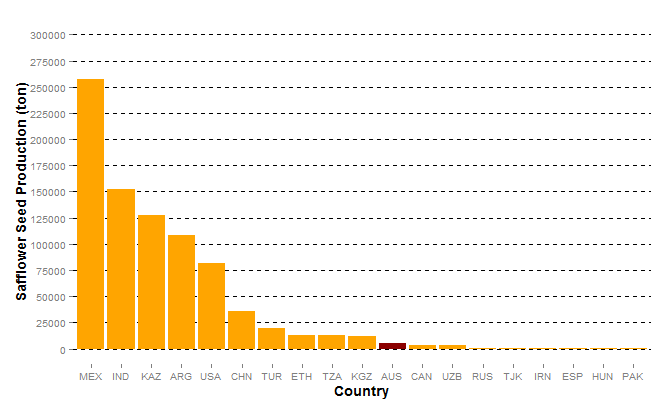


Figure 1 – Global safflower production in 2012 ordered by quantity. Australia (AUS, indicated) is ranked 11th in the world for safflower seed production at approximately 4 800 t (United Nations 2014).

comprehensive within the context of these specific species, there is very little molecular characterisation of pathways that exist across all *Asteraceae*, despite this family being the largest and most successful of the flowering plants.

## Vernalisation Response in Safflower

Of particular current interest is the coordination of flowering time in safflower, a trait that if modifiable, is likely to have enormous impact on safflower’s adaptability to climate change, increase the range of cultivatable land available for safflower crop, and total yield. Research in the late 1970s characterised both winter and spring varieties of safflower, specifically describing a low survival rate for spring safflower when planted in winter and, conversely, poor performance of winter safflower planted in the spring (Yazdi-Samadi & Zali 1979). Early research in *Arabidopsis thaliana* (*Arabidopsis*) and cereals (REFERENCES) indicated that the expression of a vernalisation response phenotype is underpinned by molecular modifications in the vernalisation response pathway, and it is this response that characterises whether a variety in a species is a ‘winter’ and ‘spring’. Therefore, it is hypothesised, and this hypothesis is supported by previous research (Johnson et al. 2006), that the ‘winter hardy’ varieties of safflower will express similar molecular and anatomical phenotypes and involve molecular pathways similar to other winter and spring varieties in other plant species (which ones?). However, it is important to note that while some safflower varieties progress rapidly to flowering following vernalisation, it is not a necessary environmental cue for safflower to flower, indicating that in safflower, the vernalisation response is *facultative* (Salisbury & Ross 1992).

While it has been documented in model plant systems such as *Arabidopsis*, as well as other agronomically important species, such as wheat and sugar beet, that vernalisation relates to crop yield in many important species, the genetic mechanism(s) of vernalisation in safflower and the effect of an extended cold treatment has not been molecularly characterised. This lack of molecular characterisation is confounded by an absence of a publicly available draft genome or transcriptomic resources. Understanding these mechanisms is essential to determining the potential of growing safflower in cooler climates, its adaptability to changing climates and the subsequent impact on yield.

* Introduce what is known about Vernalisation in other species as a bridging sentence
* *This section just ends here……..needs something else to finish it off…….. an additional paragraph or two!!!!*
* *Maybe even expand this section to talk more about the lack of genetic and molecular data for safflower which is becoming an increasingly important oil crop???*

# Vernalisation

Vernalisation, and its effect on harvest time and 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 three weeks– exposure to low but non-freezing temperature (Burn et al. 1993). In addition, in *Arabidopsis*, it has been demonstrated that the resulting time to flowering is directly proportional to the length of cold exposure (Sheldon et al. 2000) and furthermore, in numerous plant species, planting time has been optimised to increase the length of time the plant remains in the vegetative growth stage. For example, carrots (*Daucus carota*) store carbohydrates in the root organ during the vegetative growing period then, following a ‘wintering’ period where vernalisation takes place, transition to a reproductive growth state takes place, utilising this stored energy reserve for flowering and seed production (Ingram et al. 2008). While flowering is often attributed to the increased day length of spring, it is not the only factor influencing the invocation of the flowering pathway. Exposure of a seedling to an extended period of cold is also responsible for triggering and early transition from the vegetative growth the flowering. This transition is largely the result of activation of the molecular vernalisation pathway. Vernalisation is therefore an important determinant of flowering time, and flowering time is a central component for overall crop yield. Furthermore, vernalisation responses also protect delicate organs (such as those created during flowering and required for reproduction) from damage from cold exposure by restricting development until after winter has passed.

## History of Vernalisation

The necessity of a prolonged period of cold exposure to promote flowering and crop yield performance of ‘winter’ wheat cultivars was documented as early as 1857 (Klippart 1857), with this initial finding further investigated in rye and other crop species by Gassner in 1918. The term “vernalisation” was coined by Lysenko in 1928 (from the Latin *vernum*, meaning *spring*), who conducted a vast quantity of agricultural research for the Soviet Union. These early studies by Lysenko and others showed that the performance of a number of cereal cultivars increased substantially after undergoing a period of wintering when compared to their performance when planted in spring. However, Lysenko incorrectly hypothesised that the progeny of a ‘vernalised’ cereal maintained the attributes of their vernalised parents, and therefore did not need the same exposure to wintering conditions to produce the same yields(reviewed in (Chouard 1960)). In addition, in early Australian colonies, seed that had been transported over from Europe struggled under the warmer conditions and mild winters of the Australian environment, leading to widespread hunger for the early settlers when crops failed (REFERENCE – Evans 1980 – Waiting for document - PAC-10166029). It was not until nearly 100 years later through the work of William Farrar that many ‘Australian Wheats’ were developed via a traditional cross breeding approach using European wheats as the breeding lines and selecting those progeny best suited for the Australian climate (Macindoe & Brown 1968).

## The Biology of Vernalisation

Flowering in plants is a complex and intricate process. In *Arabidopsis*, , <put something in here regarding flower formation - (Zeevaart 1976). Also check out (Corbesier et al. 2007)> Grafting was originally used to demonstrate the transmissible characteristics of vernalised shoot tissue. Namely, when a vernalised shoot tip was grafted onto non-vernalised root stock, the plant will flower as if the entire plant had been exposed to the vernalisation treatment. Conversely, when a non-vernalised shoot tip was grafted onto vernalised root stock, the opposite was observed. This observation has been consistently reported for both *facultative* (where vernalisation decreases time to flowering, but is not essential for flowering), and *absolute* (where vernalisation is required to progress from vegetative growth to flowering) vernalisation sensitive species.

Contemporary research has shown that the regulatory pathways and genetic mechanisms of vernalisation are species-specific (Figure 2 below). For example, vernalisation in *Arabidopsis*,and many other dicotyledonous species (dicots), is centrally regulated via expression of the floral repressor *FLOWERING LOCUS C* (*FLC*), whereas cereals such as barley are regulated via expression of *VERNALISATION 2* (*VRN2*). Beets (*Beta vulgaris* ssp. *Vulgaris*) have a different genetic mechanism, where *FLOWERING TIME 1* (*FT1*) is responsible for regulating *FLOWERING TIME 2* (*FT2*)(Pin et al. 2010) (figure 2). Although the central mediator of vernalisation differs between plant species, the vernalisation response pathway of all vernalisation responsive plant species is mechanistically related, being epigenetic in nature, that is; environmental cues can modify gene expression between ecotype and/or cultivars of the same species, that in turn can result in the expression of a different phenotype, without any alteration to the underlying DNA sequence.

<have a read of Reeves et al 2007 for stuff about conservation of vernalisation between different species. There might be something interesting in Becker and Theissen 2003 with regard to MADS-box promoters>

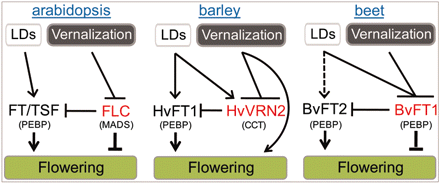


Figure 2 – The metabolic pathways central to vernalisation in *A. thaliana*, barley (*Hordeum vulgare* L.) and beet (*Beta vulgaris*) (Pin et al. 2010)

## The *Arabidopsis thaliana* vernalisation pathway

In *Arabidopsis*,FLC, a MADS-box transcription factor, mediates the transition of *Arabidopsis,* and many other dicots, from vegetative to reproductive growth (Figure 3 below). Research has shown that both genetic and epigenetic mechanisms contribute to repressing the expression of the floral repressor *FLC* during the transition from vegetative to reproductive growth (Boss et al. 2004; Finnegan et al. 2005). Namely, when *FLC* is expressed at high levels, promoted by *FRIGIDA* (*FRI*), *FRIGIDA-LIKE 1* (*FRL1*) and *FRIGIDA-LIKE 2* (*FRL2*), FLC represses the expression of *FLOWERING TIME* (*FT*), the FT homolog *TWIN SISTER OF FT* (*TSF*), and *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*), and increased FT, TSF and SOC1 levels in turn suppress expression of *LEAFY* (*LFY*) and *APETALA1* (*AP1*), the two primary promoters of floral apical meristem growth (Amasino 2004).

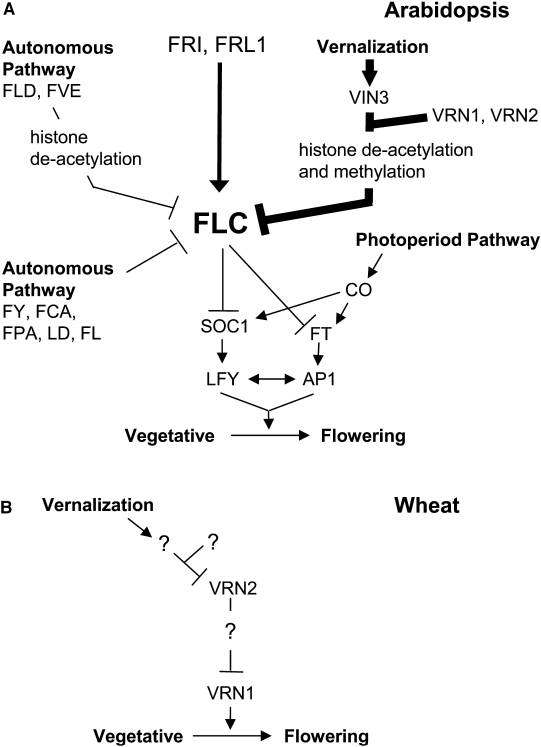


Figure 3 – The regulatory pathways involved in triggering flowering in *Arabidopsis thaliana* (Amasino 2004). FRL2 (not shown) is associated with promotion of *FLC* expression (Finnegan et al. 2005) and TSF (not shown) which is regulated in a similar fashion to *FT* (Hiraoka et al. 2013). Additional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

* I think that you could combine Figs 3 and 4 into a single, two part (3A and 3B) Figure to allow the reader to focus in on the message you are trying to convey in your text!
* The below could then be included in the same paragraph as above – currently this paragraph in disjointed by the inclusion of two Figures, one of which you have shoved right in the middle of a paragraph – very lazy!

Upon induction of FLC expression, the VERNALISATION2 (*VRN2*)/Plant Homeo domain Polycomb Repression Complex2 (PHD-PRC2) 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 conformational shape, allowing transcriptional machinery to access the *FLC* regulatory sequence and promote *FLC* expression via H3 acetylation (De Lucia et al. 2008). During vernalisation, *VERNALISATION INSENSITIVE 3 (VIN3)* levels increase. VIN3, along with *VERNALISATION 5/VIN3-LIKE* (*VEL1*) and *VERNALISATION 5* (*VRN5*), bind to the PHD-PCR2 complex to promote histone H3 deacetylation and *VRN1* and *VRN2*-directed methylation of H3K9 and H3K27. H3K27 methylation of the *FLC* locus closes the open conformation of *FLC*, blocking the transcription machinery from accessing *FLC* and thus repressing *FLC* expression .<Have a look at (Finnegan & Dennis 2007)> This epigenetic repression of *FLC* is stable and irreversible, ensuring the transition to a flowering state is permanent (Levy et al. 2002; Sung & Amasino 2004). Reduced FLC leads to increased expression of *SOC1* and *FT*, and increased SOC1 and FT levels in turn enhance the expression of the floral promoters *LFY* and *AP1 (Fig 3a)*.

<Localisation of FT expression>

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 vernalisation period (Hayama & Coupland 2004). Increased periods of day light promotes the expression of *CONSTANS* (*CO*), which in turn overrides the repressive effects of *FLC* via CO-mediated activation 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 *Arabidopsis* by down regulating *FLC* expression. *FLOWERING TIME CONTROL PROTEIN* (*FCA*) with A. thaliana *homologue of yeast RNA 3' processing factor* (*FY*), *fasciclin-like arabinogalactan* (*FLA*) and *plant specific K-homology* (*KH*)*-domain RNA-binding* (*FLK*) are independently involved with *FLC* RNA processing, *FLOWERING LOCUS D* (*FLD*) and *mammalian homologue of retinoblastoma-associated* (*FVE*) deacetylate histones at the *FLC* locus. The result is similar to the gene regulation of vernalisation, however it progresses at a much slower rate (Simpson 2004).

In *Arabidopsis*, summer ecotypes possess allelic variations but not necessarily in vernalisation genes specifically. The *Arabidopsis* 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*, which results in the early flowering of L*er*-0. Therefore, it is no longer necessary to down regulate *FLC* expression by vernalisation and the increased expression of *VRN1* and *VRN2* or long day conditions to increase the expression of *LD* (Michaels & Amasino 1999). There are many *Arabidopsis* ecotypes that can be categorised as summer and winter based on their response to vernalisation (Nordborg & Bergelson 1999).

More work needed

However, while the molecular pathways in *Arabidopsis* are well characterised, substantial differences exist in the vernalisation pathways in other plant species.

## Vernalisation pathways in cereals

In monocotyledous species (monocots), such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*) and *Brachypodium distachyon* (a model monocot), the vernalisation pathway differs greatly from *Arabidopsis*. While there are many mechanisms conserved between monocots and *Arabidopsis,* such as the Photoperiod Pathway (Griffiths et al. 2003) and FT as a regulator of flowering, the most notable difference in cereals is the absence of a known homolog for AtFLC(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) (Yan et al. 2004), along with FT, are responsible for regulation of flowering time in cereals. Before wintering, the cereal remains in a state of vegetative growth due to the presence of VRN2, which represses *FT* expression (Ream et al. 2014).. Another floral repressor, ODDSOC2 (OS2) is also present in winter cereals at high levels before vernalisation and works to maintain the cereal in a vegetative growth state (Greenup et al. 2010). Once exposed to vernalisation conditions, winter barleys increase expression of *HvVRN1* via decreased H3K27me3 and increased H3K4me3 in the *HvVRN1* locus, which opens the histone conformational shape and allows *VRN1* expression (Oliver et al. 2009). Similar to *FLC* in *Arabidopsis*, this change is stable, but in winter barley, the shape change allows access to the *HvVRN1* gene rather than restricting access. The increased levels of HvVRN1 in turn repress expression of *HvVRN2.* The decrease in HvVRN2 levels allows the expression of *FT*, which triggers the transition of the winter barley to flowering (Trevaskis et al. 2006). Exposure to vernalisation conditions and the presence of HvVRN1 also stably inhibits the expression of *OS2*, removing another block for the transition to a flowering state.

Many spring cereals contain alleles of *VRN2* that, when translated, do not suppress *FT* expression and therefore negates the requirement for vernalisation, as FT is never repressed (REFERENCE). It is also interesting to note that the response to vernalisation 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).

* 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?

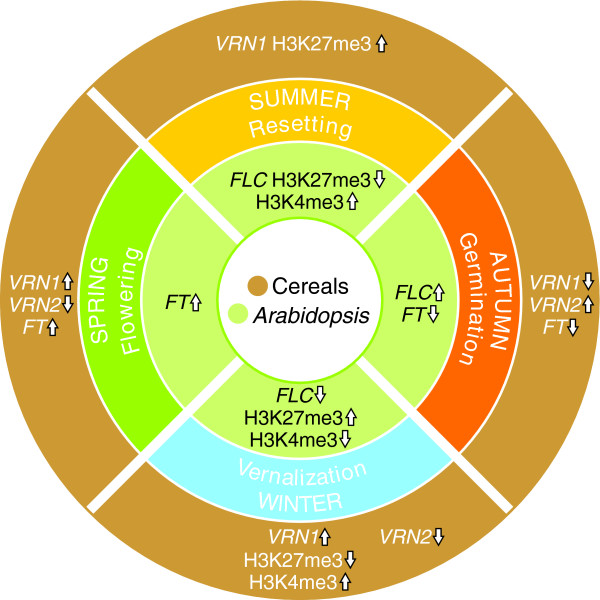


Figure 6 – Comparison of gene expression and histone modifications between vernalisation-sensitive *Arabidopsis thaliana* and “winter” cereals, the most notable difference is the absence of FLC in *A. thaliana* (Dennis & Peacock 2009). – Modify this pic somehow?

***PPD1***

***CO***

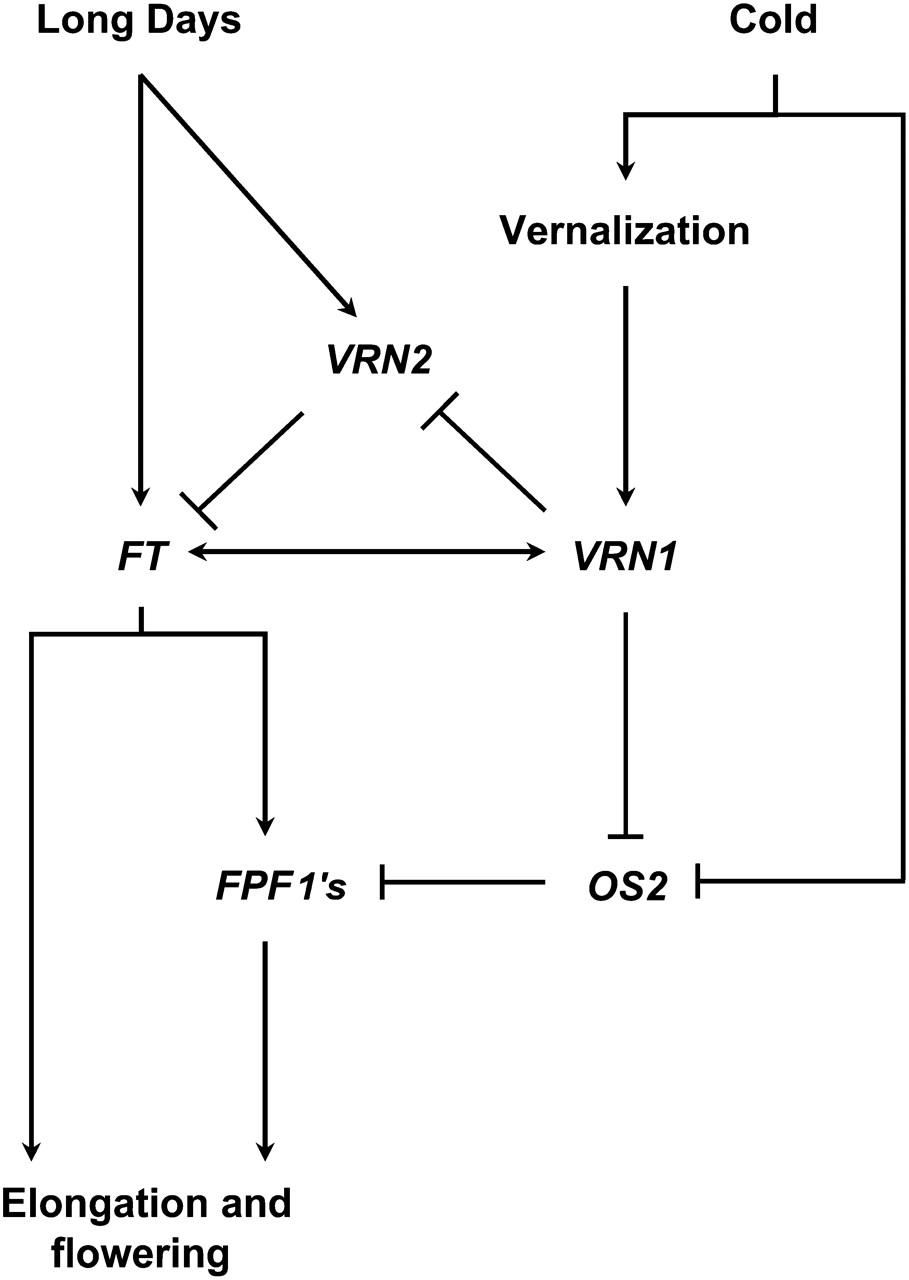


Figure 7 – Vernalisation gene regulatory pathway involved in triggering flowering in winter cereals. While the vernalisation response is different, response to day length cues contains similarities with *A. thaliana* (Modified from (Greenup et al. 2010) and (Trevaskis et al. 2007))

## Vernalisation in Sugar Beet

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 located (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 *Arabidopsis* and cereals (Figure 8 below). Rather than a single *FT* gene that triggers flowering time, as seen in *Arabidopsis*, two *FT* genes While *B. vulgaris* contains a number of genes consistent with the Photoperiod Pathway in *Arabidopsis* , the critical gene for control of transition to flowering is *BOLTING TIME CONTOL 1* (*BvBTC1*), which is   
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.

## Vernalisation in other *Asteraceae*

While much research has been undertaken with regards to the vernalisation response in *Arabidopsis*, cereals and beet, current research regarding this response to in safflower, and indeed, the *Asteraceae* family is relatively scarce, despite the *Asteraceae* being the largest flowering plant family in the world (Book at Black Mountain library by Cronquist. REFERENCE). Early research in lettuce (*Lactuca sativa*) observed that germinated seeds that were vernalised prior to planting responded by bolted up to four weeks earlier than those that were not vernalised (Warne 1947; RAPPAPORT et al. 1956). Later studies indicated that, similar to *Arabidopsis*, given the correct daylight exposure, lettuce bolts to flowering without the need for vernalisation, indicating a *facultative* vernalisation response in vernalisation sensitive lettuce species (Waycott 1995).

In Chicory (*Cichorium intybus*)*,*

Then add the Chicory data.

Hieracium? Apart from having a look through some of Anna K’s papers,

This section should not be huge, as that is the point of your entire PhD! Don’t find a paper that blows your research out of the water!

Availability of genomic resources (compositae, lettuce) will help elucidate the

Possibly allude to “winter hardy” safflower

## Determining Factors Affecting Vernalisation

There are a number of differing and redundant pathways in *Arabidopsis*, cereals and *B. Vulgaris* that can trigger flowering. While a mutation in the vernalisation pathway in a *facultative* vernalisation sensitive species may not necessarily stop the transition from vegetative growth to flowering, it will certainly prevent flowering in an *absolute* vernalisation sensitive species. For example, a mutation that removes the *CO* regulators may cause the plant to lose the requirement for lengthening light as a flowering trigger. This results in a fast transition to flowering regardless of the plant’s existing requirement for vernalisation. Recent *Arabidopsis* research indicate that higher levels of a diacyl glyceride (DAG) comprising of phosphatidylcholine (PC) and oleic acid (18:1) in the shoot apical meristem accelerates transition to flowering, and vice versa (Nakamura et al. 2014). This research also shows other DAGs containing PC binding to *FT*. This indicates that levels of certain specific lipids in plant tissues could be another contributing factor that mediates the triggering of flowering in *Arabidopsis*. However, it is unclear how lipid levels in plant tissues affect safflower’s response to vernalisation.

While safflower, *Arabidopsis* and *B. vulgaris* are all dicots, because there are so many divergences in the molecular pathways in response to vernalisation when comparing *Arabidopsis* and *B. Vulgare*, it is plausible that there are a number of factors in safflower, possibly even across the *Asteraceae* as a family, that set them apart from other dicots in the context of vernalisation. These confounding factors will have an impact on the elucidation of genes critical in the response to vernalisation especially when comparing vernalisation sensitive and vernalisation resistant safflower cultivars.

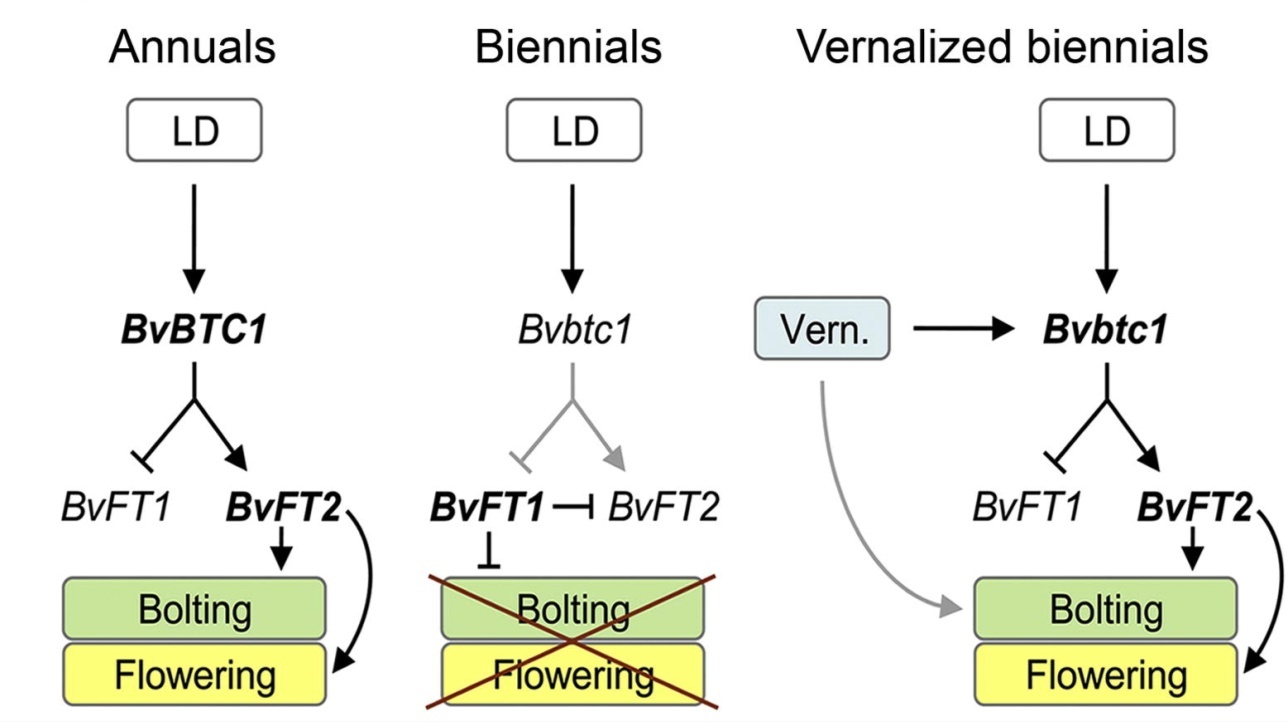


Figure 8 – Regulatory mechanisms for flowering and vernalisation in *Beta vulgaris* (Pin et al. 2012).

<segment in here linking it to the previous>

# Mapping of Pathways and Characterising Genes

## Traditional Crossing Methods

<Do I put something in here about traditional genetics i.e. cross breeding lines to segregate traits?>

## Characterisation using Forward Genetics

Forward genetics is a methodology where random genetic mutations, as either single point or clustered mutations, are introduced via chemical or radiation exposure to induce phenotypes. In diploid plants, when a germinated seed is mutated (the M1 parent), it contains random mutations on each genome in each cell, but at different positions in each genome and cells, hence these heterozygous mutations should not express any phenotype caused by genetic mutation. However, if these mutations are passed to the gametes and the plant produces seed via self-crossing, any mutations are replicated in both chromosomal copies. If this mutation is located in a transcribed or regulatory locus, it can interference with normal metabolic function, expresses as a phenotype. In each generation of mutated plants (M2, M3, M4 etc), assuming no new genetic material is introduced, these mutations remain and phenotypes continue to be expressed. DNA is sampled from plants expressing a specific phenotype and pooled together for sequencing, mapping the resulting reads against the genome. With regard to resolution of the mapping, the larger the number of individual phenotypes sampled, the clearer and stronger the signal is in in regions controlling expression of that phenotype. Because reads and mutations are mapped to specific regions of the genome, forward genetics allows mapping traits and characterisation of a genome without necessarily needing to have *a priori* knowledge of the function of that region. However, a high quality *de novo* genome, or at the very least, a quality reference of the region of interest, is required for forward genetics to be a useful and applicable approach to characterise a species <The reference for EMS and genetic screens, there are a bunch of references that can be used from here: <http://en.wikipedia.org/wiki/Genetic_screen> (Page & Grossniklaus 2002) and (Sikora et al. 2011)>

* Forward genetics – Examine the effects of point mutations and how phenotypes are expressed, mapping
  + What about in other species – Wheat?
  + Difficulties fine mapping regions of mutations relating to a phenotype without the underlying genome
  + NB: Create a diagram outlining forward genetics? Show point mutation ploidy

<for the LOVE OF GOD! CLEAN THIS UP!>

## Characterisation using Reverse Genetics

Another method to examine the underlying molecular mechanisms of a species is reverse genetics. This process examines the phenotypical effect when these genes are manipulated to either eliminate them or modifying them to make the resulting gene unusable in the system. <Zebrafish TILLING and reverse genetics - (Moens et al. 2008) > <Barley? Arabidopsis examples?> In *Nicotiana benthamiana*, the gene FATTY ACID DESATURASE 2 (NbFAD2) was knocked using an RNAi construct and two genes introduced from *Arabidopsis*, FATTY ACID ELONGASE 1 (FAE1) and acyl-CoA:diacylglycerol acyltransferase 1 (DGAT1), to manipulate the composition of the resulting leaf oil profile (Naim et al. 2012). The drawback of the reverse genetics approach is the requirement for *a priori* knowledge of expressed transcripts in the system as well as the creation of knockout constructs to successfully silence the gene in question. However, as the cost of sequencing has decreased substantially in the last decade, the creation of a useful *de novo* transcriptome is now much easier to complete, proving a wealth of candidate genes to examine with reverse genetics.

* + Any easy way of creating a diagram for reverse genetics?

# Sequencing Methodologies

The availability of genomic and transcriptomic resources has advanced biological research in leaps and bounds. Since the 1970s, genetic resources have allowed the 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 in 1975 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 in the gel corresponding to the order of nucleotides. Many significant genomes have been constructed using Sanger Sequencing, including *Caenorhabditis elegans* (The *C. Elegans* Sequencing Consortium 1998), *Arabidopsis* (*Arabidopsis* Genome Initiative 2000) and the human genome. The disadvantages of Sanger sequencing is the time and cost involved. The human genome (approximately 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, the cost of data generated by a single lane of an Illumina HiSeq 2500 sequencer is approximately $0.25 per million nucleotides. While Sanger Sequencing is the oldest, most time consuming and expensive method of sequencing per nucleotide, 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 the gene expressed. These expression profiles can be compared to identify differentially expressed genes e.g. 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 knowing 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. These rare transcripts require either *a priori* knowledge of the transcript (Adams et al. 1991) or additional molecular techniques. 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, making it difficult to identify the rare transcript due to the high signal produced by the abundant transcript.

Newer Next Generation Sequencing (NGS) technologies surpass ESTs with regard to the quantity of data produced and the accuracy of the resulting expression analysis. However, because ESTs are directly cloned from extracted mRNA and do not require assembly of fragments via algorithm, EST libraries often complement *de novo* assemblies as a means of 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 an inkjet printer (Allain et al. 2001). They 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 or sets of tiling arrays for sequencing whole genomes. Each microarray can only contain 6.5 million oligonucleotides on its surface, restricting their application 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 to sequence smaller eukaryotic genomes such as *Arabidopsis*. Similar to ESTs, the requirement for *a priori* knowledge of oligonucleotides for microarray construction and analysis decreases their efficacy for discovering *de novo* or low count transcripts, although, similar related species can be used as an adjunct.

## Next Generation Sequencing

In the last 15 years, Next Generation Sequencing(NGS) has expanded the quantity of information available to scientists to levels never before seen at an ever decreasing cost (see figure x below).

<insert figure Sequencing Costs and cost per genome (Wetterstrand 2014) >

Illumina Paired End (PE) sequencing is one of the most widespread sequencing technologies used today. It takes a DNA (or cDNA synthesised from RNA) sample and randomly fragments it. These fragments are filtered by length, oligonucleotide adaptors ligated to each end of the fragment and the entire fragment amplified. These lengths of double stranded DNA are denatured and run through a capillary tube containing oligonucleotides complementary to the adaptors. <use (Bentley et al. 2008)> A single lane on an Illumina HiSeq 2500 sequencer produces between 240 and 300 million reads of 100 bp length (approximately 24 Gnt to 30 Gnt) in under 48 hours . Even on the lower end Illumina MiSeq (a low end sequencer using Illumina technology), 20 million reads of 75 bp length (approximately 1.5 Gnt) can be produced in around the same time, but at a much lower cost to run when compared to the HiSeq 2500 (Quail et al. 2012). <link to next sentence. Something about assembly of sequences?>

An Illumina protocol for sequencing genomic DNA for the purposes of increasing the quality of an assembly is Mate Pair sequencing. This protocol does not produce long reads, but instead, produces pairs of reads with an insert length of several thousand nucleotides (Illumina 2012). While this sequencing methodology is unsuitable for transcriptomic DNA, it is swell suited for scaffolding contigs from a *de novo* genomic assembly together and orienting them in relation to one another (see “Creating a *de novo* assembly” below). In general, the error rates for PE Illumina sequencing is quite low, with a reported error rate of 0.26% for the HiSeq (that is, approximately 1 error in every 385 nt) and 0.8% for the MiSeq (1 error in every 125 nt) (Quail et al. 2012).

Another recent sequencing technology is the “Pacific Biosciences” (PacBio) protocol. <WHAT THE HELL IS IT? How does this differ from Illumina> While these reads tend to have an error rate far in excess of illumine sequencing (at approximately x / y nucleotides”, the benefit that PacBio sequencing has over Illumina is the read length. PacBio reads are on average X bp in length, #x the size of Illumina. Unfortunately, to run a single <lane> of PacBio, the cost is far greater per nucleotide than Illumina (how much? Check AGRF or others). The error rates for PacBio reads are also far in excess of Illumina reads, at around 12.86% (approximately 1 in every 7 nt) (Quail et al. 2012).

NGS technology is also being updated and revolutionised even today. One such protocol is Moleculo sequencing, which boasts the ability to sequence a single molecule of DNA, making them very useful in instances where the quantity of sample to be sequences is very low (REFERENCE!). <Chemistry/ protocol of Moleculo?>

<Something else about Ion Torrent. How is this “Important”? Something about the MinION system and “Laboratory Bench top Sequencing”> <Chemistry/ protocol of Ion Torrent?>

## Creating a *de novo* assembly

While the cost of creating next generation sequencing information has dropped dramatically in the last 15 years, only a fraction of the organisms on Earth have had their genome sequenced. This includes model organisms such as *C. elegans*, *Drosophila melanogaster* (fruit fly) (Adams et al. 2000), *Arabidopsis*, as well as other important species, such as the endangered Giant Panda (Li et al. 2010) and the platypus (Warren et al. 2008). The oldest and most common method of creating a *de novo* genome is overlapping short reads to create contigs of consensus nucleotides (Peltola et al. 1984; Huang & Madan 1999; Li et al. 2008). While this method can be effective, because each read must be compared with every other to find the best possible match. While subsequence “seeds” can be samples from a short read to find initial candidate overlaps to extend, these algorithms tend to scale quadratically. Therefore, without some way of filtering out unsuitable or duplicate reads, running the algorithm to create a *de novo* assembly soon becomes prohibitively time consuming.

An alternate method proposed for *de novo* assembly is the creation of *de brujin* graphs to traverse multiple reads. <Describe how it works (Pevzner et al. 2001)> This methodology is much more computationally efficient when compared to the traditional overlapping method. Algorithms that are based around the *De Brujin* method include Velvet (Zerbino & Birney 2008), ABySS (Simpson et al. 2009), SOAPDeNovo2 (Luo et al. 2012) and BioKanga (unpublished, http://sourceforge.net/projects/biokanga/).

Look at (Pevzner et al. 2001).

The source data that is used for a *de novo* assembly also affects the quality of the resulting assembly. In any sequencing experiment, not just *de novo* assembly, filtering reads that contain low quality scores, lengths of ambiguous nucleotides (Ns) or those reads containing substantial contamination from adaptors can decrease the time it takes for an algorithm to run simply because there are fewer reads to process. A common approach for producing a useful assembly, especially when assembling a *de novo* genome or any assembly where long contigs are important is to combine a number of different technologies in a multi-step approach (see figure X below). Short paired end reads (such as Illumina 100 bp paired end reads) are first assembled into contigs, using either *de brujin* or an overlap approach. Next, paired end mate pair reads are used to “scaffold” contigs together, that is bringing contigs together that are separated by a distance greater than the short read insert length. This orients contigs in relation to one another across a known distance, allowing gaps of unknown nucleotides but of a known length to be produced. These ambiguous regions can then be explored by using complementary primers at the known region and oriented to extend the cloned sequence across into the unknown region (REFERENCE).

Recently, long but error prone PacBio reads are now starting to complement and improve *de novo* assemblies constructed from Illumina short paired end and mate pair reads. The PacBio reads are used to bring together assembled short read contigs that cannot be brought together with mate pair scaffolding as well as a method of eliminating misassembled and chimeric contigs created when the assembly algorithm has incorrectly merged two disparate sequence segments into a single consensus contigs. In turn, the high quality short read assemblies are used as a method of correcting errors in the PacBio reads. (Bashir et al. 2012; Utturkar et al. 2014)

The problem of *de novo* assembly is so computationally complex and biologically important that “Assemblathons” are now organised to that allow bioinformaticians and software developers to bring tools together to compare algorithm efficiency and *de novo* assembly performance on reads generated from a known reference (Earl et al. 2011; Bradnam et al. 2013).

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

Regardless of whether a *de novo* assembly is genomic or transcriptomic, having no reference with which to make a comparison makes assessing the quality of a *de novo* assembly challenging. A number of techniques have had varying degrees of success addressing this problem. The “N50” value for an assembly is calculated by sorting assembly contigs in descending order and identifying the length of the contig containing 50% of the nucleotides of the entire sorted assembly (Paszkiewicz & Studholme 2010). This provides a metric where the quality of a *de novo* assembly can be assessed at a glance. However, the N50 does not provide insight regarding potential misassembly or chimeric contigs produced by the assembly algorithm (Salzberg & Yorke 2005). Because the N50 is represented as a single number, assemblies that contain a few large contigs and numerous small ones can often be misrepresented as a good quality assembly. Therefore, the N50 must not be taken as a single point of truth in an assembly, but rather as another attribute of the assembly to be used alongside others.

The most effective way to assess a transcriptome is to align 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. Aligning these 248 most common eukaryotic transcripts against the *de novo* genome can indicate assembly quality. The higher number and score of alignments of *de novo* contigs against the conserved genes, the better the quality of the assembly (Parra et al. 2007).

# Aims of this research project:

The primary aim of this research project is to investigate how vernalisation results in phenotypic variation in safflower through the examination of vernalisation sensitive and insensitive cultivars*.* There are a number of methods that will be utilised 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 safflower tissues will be analysed to determine their candidature as vernalisation genes, as well as genes originating from other plant species sharing sequenced similar to the *de novo* transcripts. The *in silico* candidate genes will be verified or refuted using molecular biology techniques. The results of this research will influence traditional breeding techniques and the genetic manipulation of safflower cultivars with the goal of diversifying the regions of Australia where safflower can be grown.

# References

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

Adams, M.D. et al., 2000. The Genome Sequence of Drosophila melanogaster. *Science*, 287 (5461 ), pp.2185–2195. Available at: http://www.sciencemag.org/content/287/5461/2185.abstract.

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.

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 408(6814), pp.796–815. Available at: http://dx.doi.org/10.1038/35048692.

Bashir, A. et al., 2012. A hybrid approach for the automated finishing of bacterial genomes. *Nat Biotech*, 30(7), pp.701–707. Available at: http://dx.doi.org/10.1038/nbt.2288.

Bentley, D.R. et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–59. Available at: http://dx.doi.org/10.1038/nature07517.

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.

Bradnam, K. et al., 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, 2(1), p.10. Available at: http://www.gigasciencejournal.com/content/2/1/10.

Burn, J.E. et al., 1993. DNA methylation, vernalization, and the initiation of flowering. *Proceedings of the National Academy of Sciences*, 90(1), pp.287–291. Available at: http://www.pnas.org/content/90/1/287.abstract.

Cao, S. et al., 2013. A large and functionally diverse family of Fad2 genes in safflower (Carthamus tinctorius L.). *BMC Plant Biology*, 13(1), p.5. Available at: http://www.biomedcentral.com/1471-2229/13/5.

Chempro Technovation 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.

Corbesier, L. et al., 2007. FT Protein Movement Contributes to Long-Distance Signaling in Floral Induction of Arabidopsis. *Science*, 316(5827), pp.1030–1033. Available at: http://www.sciencemag.org/content/316/5827/1030.abstract.

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.

Earl, D.A. et al., 2011. Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Research*. Available at: http://genome.cshlp.org/content/early/2011/09/16/gr.126599.111.abstract.

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.

Finnegan, E.J. & Dennis, E.S., 2007. Vernalization-Induced Trimethylation of Histone H3 Lysine 27 at FLC Is Not Maintained in Mitotically Quiescent Cells. *Current Biology*, 17(22), pp.1978–1983. Available at: http://www.cell.com/current-biology/abstract/S0960-9822(07)02087-8.

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.

Huang, X. & Madan, A., 1999. CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9(9), pp.868–877. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC310812/.

Illumina, 2012. Data Processing of Nextera Mate Pair Reads on Illumina Sequening Platforms. Available at: http://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote\_nextera\_matepair\_data\_processing.pdf [Accessed January 4, 2015].

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.

Klippart, J.H., 1857. An essay on the origin, growth, diseases, varieties etc of the wheat plant. *Annual Report of the Ohio State Board of Agriculture*, pp.562–720.

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.

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.

Koltunow, A.M.G., Johnson, S.D. & Okada, T., 2011. Apomixis in hawkweed: Mendel’s experimental nemesis. *Journal of Experimental Botany*, 62(5), pp.1699–1707. Available at: http://jxb.oxfordjournals.org/content/62/5/1699.abstract.

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.

Li, R. et al., 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics*, 24(5), pp.713–714. Available at: http://bioinformatics.oxfordjournals.org/content/24/5/713.abstract.

Li, R. et al., 2010. The sequence and de novo assembly of the giant panda genome. *Nature*, 463(7279), pp.311–317.

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.

Luo, R. et al., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1), p.18. Available at: http://www.gigasciencejournal.com/content/1/1/18.

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.

Moens, C.B. et al., 2008. Reverse genetics in zebrafish by TILLING. *Briefings in Functional Genomics and Proteomics*, 7(6), pp.454–459. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2899843/.

Naim, F. et al., 2012. Advanced Engineering of Lipid Metabolism in Nicotiana benthamiana Using a Draft Genome and the V2 Viral Silencing-Suppressor Protein. *PLoS ONE*, 7(12), p.e52717. Available at: http://dx.doi.org/10.1371/journal.pone.0052717.

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.amjbot.org/content/86/4/470.abstract.

Oliver, S.N. et al., 2009. Vernalization-induced flowering in cereals is associated with changes in histone methylation at the VERNALIZATION1 gene. *Proceedings of the National Academy of Sciences*, 106(20), pp.8386–8391. Available at: http://www.pnas.org/content/106/20/8386.abstract.

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.

Page, D.R. & Grossniklaus, U., 2002. The art and design of genetic screens: Arabidopsis thaliana. *Nat Rev Genet*, 3(2), pp.124–136. Available at: http://dx.doi.org/10.1038/nrg730.

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.

Paszkiewicz, K. & Studholme, D.J., 2010. De novo assembly of short sequence reads. *Briefings in Bioinformatics*, 11(5), pp.457–472. Available at: http://bib.oxfordjournals.org/content/11/5/457.abstract.

Peltola, H., Söderlund, H. & Ukkonen, E., 1984. SEQAID: a DNA sequence assembling program based on a mathematical model. *Nucleic Acids Research*, 12(1 Pt 1), pp.307–321. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC321006/.

Pevzner, P.A., Tang, H. & Waterman, M.S., 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 98(17), pp.9748–9753. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC55524/.

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.

Quail, M. et al., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1), p.341. Available at: http://www.biomedcentral.com/1471-2164/13/341.

RAPPAPORT, L., WITTWER, S.H. & TUKEY, H.B., 1956. Seed Vernalization and Flowering in Lettuce (Lactuca sativa). *Nature*, 178(4523), p.51. Available at: http://dx.doi.org/10.1038/178051a0.

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.

Salzberg, S.L. & Yorke, J.A., 2005. Beware of mis-assembled genomes. *Bioinformatics*, 21(24), pp.4320–4321. Available at: http://bioinformatics.oxfordjournals.org/content/21/24/4320.short.

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.

Sikora, P. et al., 2011. Mutagenesis as a tool in Plant Genetics, Functional Genomics and Breeding. *International Journal of Plant Genomics*, p.13.

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.

Simpson, J.T. et al., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research*, 19(6), pp.1117–1123. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2694472/.

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.

The C. Elegans Sequencing Consortium, 1998. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282(5396), pp.2012–2018. Available at: http://www.sciencemag.org/content/282/5396/2012.abstract.

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.

Utturkar, S.M. et al., 2014. Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics*, 30(19), pp.2709–2716. Available at: http://bioinformatics.oxfordjournals.org/content/30/19/2709.abstract.

Warne, L.G.G., 1947. Vernalization of Lettuce. *Nature*, 159(4027), pp.31–32.

Warren, W.C. et al., 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature*, 453(7192), pp.175–183. Available at: http://dx.doi.org/10.1038/nature06936.

Waycott, W., 1995. Photoperiodic Response of Genetically Diverse Lettuce Accessions. *Journal of the American Society for Horticultural Science*, 120(3), pp.460–467. Available at: http://journal.ashspublications.org/content/120/3/460.abstract.

Wetterstrand, K., 2014. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: http://www.genome.gov/sequencingcosts/ [Accessed December 27, 2014].

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.

Zeevaart, J.A.D., 1976. Physiology of Flower Formation. *Annual Review of Plant Physiology*, 27(1), pp.321–348. Available at: http://dx.doi.org/10.1146/annurev.pp.27.060176.001541.

Zerbino, D.R. & Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5), pp.821–829. Available at: http://genome.cshlp.org/content/18/5/821.abstract.

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