DOCTORAL THESIS

The Molecular Characterisation of the Vernalisation Response in Safflower via the Development of Genomic and Transcriptomic Resources

Author:
Darren CULLERNE
BAppSci/BIT
MBiotech(*Research*)

Supervisors:
Dr. Andy EAMENS
Dr. Craig WOOD
Dr. Ben TREVASKIS
Prof. Christopher GROF

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy (Biological Sciences)

in the

Faculty of Science and Information Technology
School of Environmental and Life Sciences

May 3, 2017

Statements of Collaboration and Originality

I, Darren CULLERNE, declare that:

(a) This thesis, titled 'The Molecular Characterisation of the Vernalisation Response in Safflower via the Development of Genomic and Transcriptomic Resources' contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository**, subject to the provisions of the Copyright Act 1968.
**Unless an Embargo has been approved for a determined period.

(b) The work embodied in this thesis has been conducted in collaboration with and carried out at CSIRO Black Mountain in Canberra, in association with the University of Newcastle. As part of this collaboration, I have undertaken the research contained within this Thesis using CSIRO facilities under Dr Craig Wood in the Safflower Engineering laboratories.

Signed: _____

Date: _____

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Craig Wood, Dr Ben Trevaskis, Dr Andy Eamens and Professor Chris Grof. Each of these people provided a unique depth of knowledge and expertise in their guiding contributions to this project.

I would like to acknowledge: The University of Newcastle, the Grains Research and Development Corporation and the CSIRO's Office of the Chief Executive for providing both operating funds and scholarship stipends. My involvement in this project would not have been possible without their financial support.

The bioinformatics community at CSIRO. In particular, my conversations with Dr Stuart Stephen, Mr Andrew Spriggs and Dr Jen Taylor, many of which were facilitated by the consumption of coffee. They proved to be invaluable for the correct planning and analysis of the countless samples and data files processed as part of this project.

The many postdoctoral scientists and technicians in the CSIRO Plant and Oil Engineering and the CSIRO Crop Genomics groups. Their suggestions of protocols, troubleshooting help, molecular biology techniques and assistance in getting my hands dirty on the bench and in the glasshouses is much appreciated.

The other PhD students, both within and external to CSIRO, some who have already graduated and others who yet to complete their projects. Their camaraderie and support helped me through some rough times during this project. I hope I have helped them in the same way. I would like to particularly thank Kyle Reynolds for his mental support, baked goods and coffee breaks throughout this project.

My family and friends for their ongoing support of my partner and I. Whether the support was financial, psychological, fermented or nutritional, I could not have accomplished this project without them.

Lastly, but most importantly, I'd like to thank my partner Michelle for support and encouragement throughout my studies and during this project in particular. Her endless patience and unwavering belief in me were crucial to getting me over the finishing line.

To my grandparents, Eileen Dotridge Comans, Thomas Joseph Bailey and Monica Ann Bailey and my great uncle Edmund Theopolis Cullerne. You have uniquely influenced, supported, and inspired me to excel.

Contents

List of Figures

List of Tables

Abbreviations

AGRF	**A**ustralian **G**enome **R**esearch **F**acility
Arabidopsis	*Arabidopsis thaliana*
AraTha	***Ara**bidopsis **tha**liana*
At	***A**rabidopsis **t**haliana*
BLAST	**B**asic **L**ocal **A**lgorithm **S**earch **T**ool (Software)
BLASTN	**BLAST N**ucleotide (Software)
BLASTP	**BLAST P**rotein (Software)
BUSCO	**B**enchmarking **U**niversal **S**ingle-**C**opy **O**rthologs (Software)
bp	**b**ase **p**air
Bv	***B**eta **v**ulgaris*
CarTin	***Car**thamus **tin**ctorius*
CEGMA	**C**ore **E**ukaryotic **G**enes **M**apping **A**pproach (Software)
ChrLav	***Chr**ysanthemum **lav**andulifolium*
ChrMor	***Chr**ysanthemum **mor**ifolium*
Ci	***C**hicory **i**ntybus*
cM	**C**enti**m**organs
CSIRO	**C**ommonwealth **S**cientific and **I**ndustrial **R**esearch **O**rganisation
Ct	***Car**thamus **tin**ctorius*
CTAB	**C**etyl **t**rimethyl **a**mmonium **b**romide
DArT	**D**iversity **Ar**rays **T**echnology
DEPC	**Die**thyl**p**yro**c**arbonate
EDTA	**E**thylene**d**iamine**t**etraacetic **a**cid
Eg	*Eustoma **g**randiflorum*
EGTA	**E**thylene **g**lycol-bis(β-aminoethyl ether)-N,N,N′,N′-**t**etraacetic **a**cid
E/O	**E**ocene / **O**ligocene boundary
EST	**E**xpressed **S**equence **T**ag
g	**g**ravitational force
Gbp	**G**iga**b**ase **p**air (1 Gbp = 1,000,000,000 bp)
GLA	**g**amma **l**inolenic **a**cid
GM	**G**enetic **M**odification
GRDC	**G**rains **R**esearch **D**evelopment **C**orporation
ha	**h**ect**a**re
HorVar	***Hor**deum **vul**gare*
Hv	***H**ordeum **v**ulgare*
MP	**M**ate **p**air (reads)
MSA	**M**ultiple **s**equence **a**lignment
Mt	***M**edicago **t**runcatula*
M. truncatula	*Medicago truncatula*

mya	**M**illions of years (**A**nnus)
NCBI	**N**ational **C**enter for **B**iotechnology **I**nformation
NEB	**N**uclear **E**xtraction **B**uffer
NGS	**N**ext **g**eneration **s**equencing
$OD_{x/y}$	**O**ptical **d**ensity (absorbance) at wavelength x and y
PacBio	**Pac**ific **Bio**sciences (sequencing technology)
PCR	**p**olymerase **c**hain **r**eaction
PE	**P**aired **e**nd (reads)
PHD-PRC2	**P**lant **H**omeo-**D**omain **P**olycomb **R**epression **C**omplex **2**
PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
PVP	**p**oly**v**inyl**p**yrrolidinone
RIN	**R**NA **in**tegrity score
RPKM	**R**eads **p**er **k**ilobase of transcript per **m**illion mapped reads
RT-qPCR	**R**everse **t**transcriptase **q**uantitative **p**olymerase **c**hain **r**eaction
SAM	**S**hoot **a**pical **m**eristem
SCUBAT	**S**caffolding **C**ontigs **U**sing **B**LAST-like **A**lignment **T**ool
SDS	**S**odium **d**odecyl **s**ulfate
SHO	**S**uper **h**igh **o**leic
SNP	**s**ingle **n**ucleotide **p**olymorphism
TAIR	**T**he **A**rabidopsis *I*nformation **R**esource
Tris-HCl	**Tris**(hydroxymethyl)aminomethane hydrochloride
UTR	**Un**translated **r**egion

Gene and Protein Abbreviations

AP1	APETALA 1
CAL	CAULIFLOWER-A
BTC1	BOLTING TIME CONTROL 1
FAD2	FATTY ACID DESATURASE 2
FD	BASIC-LEUCINE ZIPPER (BZIP) TRANSCRIPTION FACTOR
FL	FLC-LIKE (*Beta vulgaris*)
FLC	FLOWERING LOCUS C
FLCL	FLC-LIKE (*Eustoma* spp.)
FL1	FLC-LIKE 1 (*Chicory intybus*)
FRI	FRIGIDA
FT	FLOWERING LOCUS T
FTL	FT-LIKE
LFY	LEAFY
MAF1-5	MADS AFFECTING FLOWERING 1-5
OS2	ODDSOC2
PEP1	PERPETUAL FLOWERING 1
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
SOC1L	SOC1-LIKE
VIN3	VERNALISATION INSENSITIVE 3
VRN1-2	VERNALISATION 1-2

Nomenclature 1 Italicised represents the DNA locus or RNA transcript of the gene. Non-italicised represents the encoded protein.

Nomenclature 2 A lower case gene or encoded protein represents a mutant or recessive allele of that gene or protein.

Nomenclature 3 A two letter abbreviation before a gene represents the species where it is found, e.g. *At* = *Arabidopsis*.

Abstract

The Molecular Characterisation of the Vernalisation Response in Safflower via the Development of Genomic and Transcriptomic Resources

by Darren CULLERNE

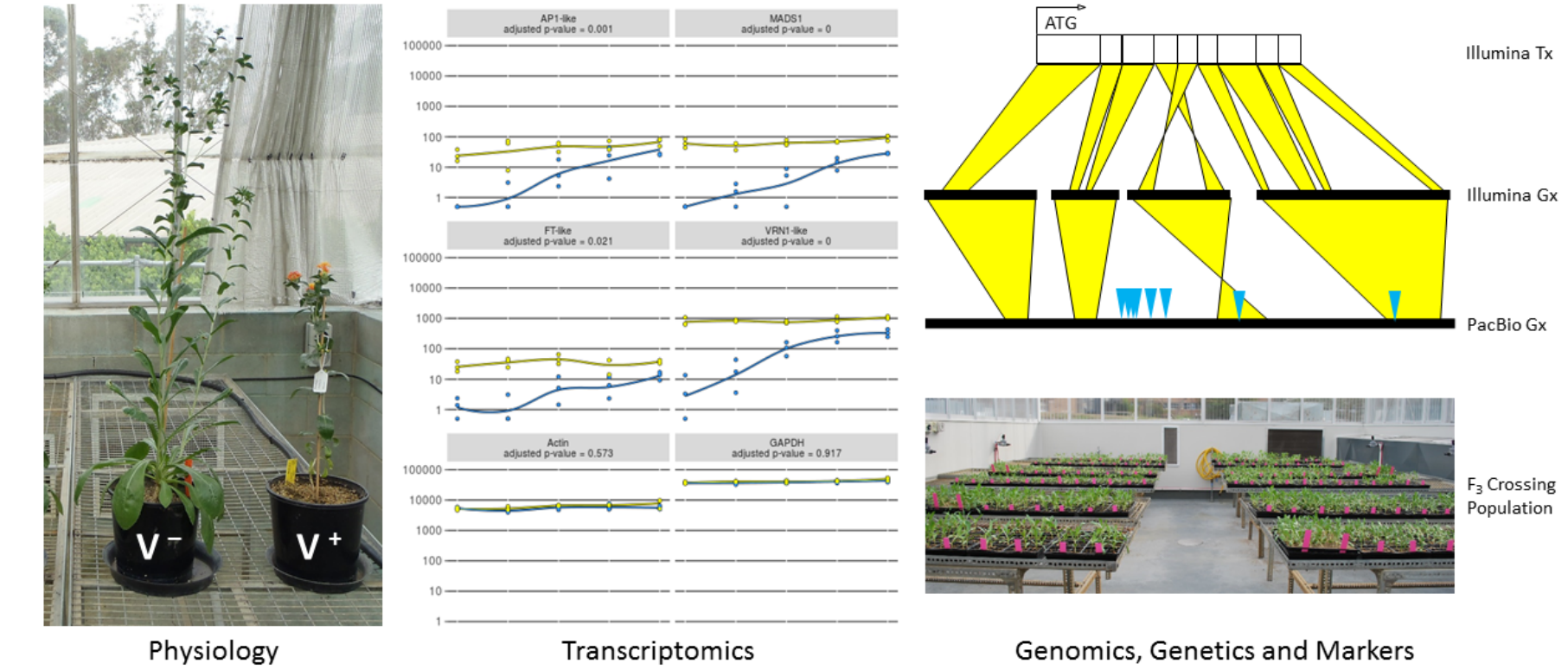
Safflower (*Carthamus tinctorius*) is an oilseed grown globally. There is an interest in modifying safflower to cope with climate change and to adapt to new agronomic trends. While almost all cropped safflower are spring varieties, a number of wild safflower varieties are noted as 'winter hardy'. These display characteristics found in plants that respond to vernalisation (an extended period of non-freezing cold). Because flowering traits, including the vernalisation response, are linked to yield and adaptability to climate, this PhD project sought to understand the vernalisation response in safflower, as a component of flowering time.

The vernalisation response in 'winter hardy' and spring safflower was investigated. It was confirmed that winter safflower does respond to vernalisation conditions, similar to other plant species. The vernalisation response in winter safflower is saturated after approximately 2 weeks exposure to vernalisation conditions. It is inheritable, epigenetic and appears to be dependent on a single recessive allele, as shown by segregation ratios in a crossing population created from winter and spring parents.

Two approaches were developed to characterise this vernalisation response. Firstly, RNA sequencing (RNA-Seq) was performed on RNA extracted from winter and spring safflower before, during and after exposure to vernalisation conditions. Differential expression analyses on the resulting RNA-Seq datasets tentatively identified four genes as directing functional roles in the vernalisation response of safflower: *APETALA 1-LIKE* (*CtAP1-LIKE*), *MADS-BOX DOMAIN CONTAINING 1* (*CtMADS1*), *FLOWERING LOCUS T-LIKE* (*CtFT-LIKE*) and *VERNALISATION 1-LIKE* (*CtVRN1-LIKE*). This analysis also identified 33 additional gene products (annotated transcripts or transcripts of unknown function) as candidates for further experimental investigation.

In addition to the transcriptomic data, genomic resources were developed to further characterise the molecular basis of the vernalisation response. A high quality *de novo* assembly was constructed using Illumina reads from spring safflower, covering approximately 80% of the estimated 1,400,000,000 base pair (1.4 Gbp) spring safflower genome. Using this draft genome in combination with F_3 crossing families and a genetic marker approach, 27 genetic markers for vernalisation were identified. Furthermore, these markers were mapped to a recent genetic map of safflower (Bowers 2016), clustering in close proximity to one another on chromosome 8. A single differentially expressed transcript, identified in the transcriptomic analyses, was located on the same chromosome. However, the transcript of interest was mapped to a chromosome 8 position some distance away from the identified markers.

These high quality transcriptomic and genomic resources were used to identify the molecular basis for vernalisation in safflower. The investigative approaches developed in this project can also be utilised to characterise the molecular mechanisms of other traits in safflower.



Graphical abstract of this PhD project. The first panel shows winter safflower, unvernalised and vernalised. The second panel shows four characterised transcripts that are differentially expressed after vernalisation of winter safflower, with two non-differentially expressed transcripts at the bottom. Blue are from winter safflower, yellow are from spring safflower The third panel shows a transcript mapping to multiple Illumina genomic contigs, which in turn map to a single PacBio contig. The bottom of the third panel shows the F_3 crossing population.