

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/77863>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

**Delayed Bolting in Rocket for Improved Quality and  
Greater Sustainability**

**Jemma Louise Taylor**

**PhD – Plant and Environmental Sciences**

**School of Life Sciences, University of Warwick**

**September 2015**

# Table of Contents

A list of tables and figures .....	vii
Acknowledgements.....	xi
Declaration.....	xii
Summary.....	xiii
Abbreviations.....	xiv
Chapter 1. Introduction.....	1
1.1. Project background.....	1
1.2. Overall aim.....	2
1.3. Objectives.....	2
1.4. Introduction to rocket.....	2
1.4.1. Overview.....	2
1.4.2. Phylogeny.....	3
1.4.3. Cultivation.....	4
1.4.4. Taste and health.....	5
1.4.5. Breeding.....	6
1.4.6. Bolting.....	6
1.5. Introduction to flowering.....	7
1.5.1. Overview.....	7
1.5.2. Photoperiodic pathway.....	8
1.5.3. Autonomous pathway.....	10
1.5.4. Vernalization pathway.....	10
1.5.5. Age dependant pathway.....	11
1.5.6. Ambient temperature pathway.....	12
1.5.7. Gibberellic acid pathway.....	12
1.5.8. Role of flowering pathway genes in crop species.....	13
1.6. Project strategy.....	14
1.6.1. Understanding the flowering pathway in rocket.....	14
1.6.2. Use of mutagenesis to delay the onset of flowering in rocket.....	16
1.7. Elsom's Seeds Ltd and CASE award.....	17
Chapter 2. Materials and Methods.....	19
2.1. Materials.....	19
2.1.1. Arabidopsis.....	19
2.1.2. Rocket ( <i>Diplotaxis tenuifolia</i> ).....	19
2.1.3. Scoring plants for bolting.....	20
2.1.4. Rocket pollination techniques.....	21
2.1.5. Seed threshing.....	23

2.1.6.	Rocket ethyl methanesulphonate (EMS) mutagenesis.....	23
2.1.7.	Growth conditions.....	26
2.2.	Standard laboratory methods.....	28
2.2.1.	DNA extraction.....	28
2.2.2.	RNA extraction and cDNA synthesis .....	28
2.2.3.	Primer design .....	29
2.2.4.	Polymerase Chain Reaction (PCR).....	29
2.2.5.	PCR product visualisation and purification .....	29
2.2.6.	DNA sequencing.....	30
2.2.7.	GeneRacer™.....	30
2.2.8.	Molecular cloning .....	30
2.2.9.	Functional complementation of Arabidopsis mutants .....	32
2.2.10.	Real time PCR gene expression analysis .....	33
Chapter 3.	Flowering Time Gene Isolation .....	35
3.1.	Introduction .....	35
3.1.1.	CONSTANS .....	35
3.1.2.	GIGANTEA.....	36
3.1.3.	FLOWERING LOCUS C .....	36
3.1.4.	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 .....	37
3.1.5.	FLOWERING LOCUS T .....	37
3.2.	Methods.....	38
3.2.1.	Gene isolation .....	38
3.2.2.	Functional complementation of Arabidopsis mutants .....	39
3.2.3.	Diurnal time course experiment.....	40
3.2.4.	Developmental time course.....	41
3.2.5.	Gene expression analysis using real time PCR.....	41
3.3.	Results .....	41
3.3.1.	CONSTANS (CO) .....	41
3.3.2.	GIGANTEA (GI).....	49
3.3.3.	FLOWERING LOCUS C (FLC) .....	52
3.3.4.	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1).....	60
3.3.5.	FLOWERING LOCUS T (FT) .....	64
3.4.	Discussion .....	86
3.4.1.	CONSTANS .....	86
3.4.2.	GIGANTEA.....	87
3.4.3.	FLOWERING LOCUS C .....	89
3.4.4.	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 .....	90

3.4.5.	FLOWERING LOCUS T and FT-like genes.....	91
Chapter 4.	Investigating the Photoperiodic and Vernalization Responses in Rocket .....	94
4.1.	Introduction .....	94
4.1.1.	Vernalization.....	94
4.1.2.	Photoperiodic response .....	95
4.2.	Methods.....	96
4.2.1.	Rocket seed .....	96
4.2.2.	Vernalization experiments .....	96
4.2.3.	Photoperiodic experiments.....	98
4.2.4.	Gene expression analysis using real time PCR.....	99
4.3.	Results .....	100
4.3.1.	Vernalization experiments .....	100
4.3.2.	Photoperiodic experiments.....	106
4.4.	Discussion .....	108
4.4.1.	Vernalization experiments .....	108
4.4.2.	Photoperiodic experiments.....	110
Chapter 5.	Production and Identification of Late Bolting Lines in Rocket.....	112
5.1.	Introduction .....	112
5.1.1.	Mutagenesis by ethyl methanesulphonate .....	112
5.1.2.	Utilising EMS to identify desired phenotypes .....	112
5.2.	Methods.....	114
5.2.1.	EMS treatment .....	114
5.2.2.	Whole genome sequencing .....	115
5.2.3.	Pipeline for genome assembly and SNP identification.....	115
5.3.	Results .....	116
5.3.1.	EMS mutagenesis.....	116
5.3.2.	Identification of lines suitable for whole genome sequencing.....	126
5.3.3.	SNP identification.....	130
5.4.	Discussion .....	131
Chapter 6.	Targeted Mutagenesis .....	134
6.1.	Introduction .....	134
6.1.1.	Targeted mutagenesis methods.....	135
6.1.2.	Optimisation of CRISPR/Cas9 methods.....	141
6.2.	Methods.....	142
6.2.1.	CRISPR design and construction.....	142
6.2.2.	Stable transformation .....	146
6.2.3.	Transient transformation.....	148

6.3.	Results .....	152
6.3.1.	Analysis of transformed plants .....	152
6.3.2.	Transient transformation .....	166
6.4.	Discussion .....	172
6.4.1.	Production of CRISPR transformed plants .....	172
6.4.2.	Transient expression in protoplasts.....	177
Chapter 7.	General Discussion .....	180
7.1.	Isolation and characterisation of Arabidopsis flowering time gene homologues in rocket 180	
7.2.	Understanding the vernalization and photoperiodic requirements of rocket .....	182
7.3.	Identification of LB rocket lines from an EMS population .....	185
7.4.	Targeted mutagenesis of <i>FT</i> using CRISPR/Cas9 in rocket and Arabidopsis .....	187
7.5.	Conclusions .....	189
7.6.	Future Work .....	192
7.7.	Elsoms Seeds Ltd and CASE award .....	194
References.....		195
Appendix A.....		216
A.1.	Primers used to isolate gene sequences.....	216
A.1.1.	CONSTANS .....	216
A.1.2.	GIGANTEA.....	216
A.1.3.	FLOWERING LOCUS C .....	218
A.1.4.	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 .....	219
A.1.5.	FLOWERING LOCUS T .....	220
A.1.6.	TWIN SISTER OF FT .....	221
A.2.	Other Primers .....	222
Appendix B.....		224
B.1.1.	Gene and protein sequences of isolated flowering time genes in rocket .....	224
Appendix C.....		238
C.1.	SNPs identified in whole genome sequencing .....	238
C.1.1.	SNP 1) 85295 C1_0 C2_18 C1_14 C2_0 A/G .....	238
C.1.2.	SNP 2) 76246 C1_8 C2_0 C1_0 C2_9 A/G .....	238
C.1.3.	SNP 3) 77982 C1_0 C2_8 C1_12 C2_0 C/T.....	239
C.1.4.	SNP 4) 141907 C1_0 C2_8 C1_8 C2_0 A/G .....	239
C.1.5.	SNP 5) 63435 C1_0 C2_9 C1_13 C2_0 C/T.....	240
C.1.6.	SNP 6) 109293 Q1_72 Q2_0 , A/G rank_0.81507, .....	240
Appendix D.....		241
D.1.	Protoplast isolation results .....	241

D.2. Primers used in targeted mutagenesis experiments.....	242
D.2.1. CRISPR <i>FT</i> target primers.....	242
D.2.2. Primers for the isolation of <i>AtFT</i> .....	242
D.2.3. Primer pairs used in PCR for cloning of shorter <i>AtFT</i> fragments.....	243

## A list of tables and figures

Figure 1.1 – A <i>Diplotaxis tenuifolia</i> plant showing typical dark green and lobed leaves .....	3
Figure 1.2 – Phylogenetic tree of rocket (Sayers <i>et al.</i> , 2011).....	4
Figure 1.3 – Lawnmower style harvesting machine used for harvesting leafy salad crops including rocket. ....	5
Figure 1.4 – Flowering time regulatory pathways in Arabidopsis.....	8
Figure 1.5 – Role of key photoperiod genes <i>GI</i> , <i>FKF1</i> , <i>CDF</i> and <i>CO</i> in Arabidopsis under long day and short day conditions. ....	9
Figure 2.1 – Scoring rocket plants for bolting. ....	21
Figure 2.2 – Bagged plant with flies to aid pollination.....	22
Figure 2.3 – Bud pollination of <i>Diplotaxis tenuifolia</i> .....	23
Figure 3.1 – pB2GW7 plasmid used as binary vector in complementation experiments.....	40
Figure 3.2 – Structure of <i>DtCO</i> gene from <i>Diplotaxis tenuifolia</i> . ....	42
Figure 3.3 – <i>DtCO</i> protein BLASTP result. <i>DtCO</i> protein (Query) has been aligned to <i>B. napus</i> CO homologue (sbjct). ....	42
Figure 3.4 – Flowering time for T <sub>1</sub> generation of <i>AtLer:DtCO</i> overexpression lines. ....	43
Figure 3.5 – Flowering time and transgene presence for T <sub>1</sub> generation of <i>co-2:DtCO</i> complementation lines. ....	44
Figure 3.6 – Flowering time data and transgene amplification for <i>AtLer:DtCO</i> 11 T <sub>2</sub> line....	45
Figure 3.7 – Flowering time data for <i>co-2:DtCO</i> T <sub>2</sub> lines.....	46
Figure 3.8 – Transgene amplification in complementation line <i>co-2:DtCO</i> T <sub>2</sub> . ....	47
Figure 3.9 – Diurnal Time Course of <i>DtCO</i> . ....	48
Figure 3.10 – Expression of <i>DtCO</i> across development.....	48
Figure 3.11 – Structure of <i>DtGI</i> gene CDS from <i>Diplotaxis tenuifolia</i> . ....	49
Figure 3.12 – DtGI protein BLASTP result.....	50
Figure 3.13 – Diurnal Time Course of <i>DtGI</i> .....	51
Figure 3.14 – Expression of <i>DtGI</i> across development. ....	52
Figure 3.15 – Structure of <i>DtFLC</i> gene CDS from <i>Diplotaxis tenuifolia</i> .....	53
Figure 3.16 – DtFLC protein BLASTP result. DtFLC protein (Query) has been aligned to <i>Sinapis alba</i> FLC homologue (sbjct). ....	53
Figure 3.17 – Flowering time of T <sub>1</sub> generation of <i>AtCol0:DtFLC</i> and <i>FRI<sup>+</sup>flc3:DtFLC</i> transformed plants.....	54
Figure 3.18 – Flowering time data for <i>AtCol0:DtFLC</i> T <sub>2</sub> lines. ....	56
Figure 3.19 – Flowering time data for <i>FRI<sup>+</sup>flc3:DtFLC</i> T <sub>2</sub> lines. ....	57
Figure 3.20 – Flowering time and transgene expression of T <sub>2</sub> generation plants from lines <i>AtCol0:DtFLC</i> 12, <i>FRI<sup>+</sup>flc3:DtFLC</i> 5 and <i>FRI<sup>+</sup>flc3:DtFLC</i> 23.....	58
Figure 3.21 – Diurnal Time Course of <i>DtFLC</i> . ....	59
Figure 3.22 – Expression of <i>DtFLC</i> across development.....	60
Figure 3.23 – Structure of <i>SOC1</i> gene CDS from <i>Diplotaxis tenuifolia</i> . ....	61
Figure 3.24 – DtSOC1 protein BLASTP result. ....	61
Figure 3.25 – Flowering time data by rosette leaf number at 1cm bolt for T <sub>1</sub> generation of <i>AtCol0:DtSOC1</i> overexpression lines.....	62

Figure 3.26 – Flowering time data by rosette leaf number at 1cm bolt for T <sub>1</sub> generation of <i>soc1-2:DtSOC1</i> expression lines. ....	62
Figure 3.27 – Transgene amplification in complementation line <i>AtCol0:DtSOC1</i> and <i>soc1-2:DtSOC1</i> .....	63
Figure 3.28 – Diurnal Time Course of <i>DtSOC1</i> . ....	63
Figure 3.29 – Expression of <i>DtSOC1</i> across development.....	64
Figure 3.30 – Structure of <i>DtFT</i> gene CDS from <i>Diplotaxis tenuifolia</i> . ....	65
Figure 3.31 – <i>DtFT</i> alignment result from blastn. ....	65
Figure 3.32 – DtFT protein BLASTP result. ....	66
Figure 3.33 – Structure of <i>DtTSF</i> gene CDS from <i>Diplotaxis tenuifolia</i> . ....	67
Figure 3.34 – DtTSF protein BLASTP result. ....	68
Figure 3.35 – Sequence distances and phylogenetic tree comparing DtFT and DtTSF with AtFT, AtTSF, AtBFT and AtTFL.....	68
Figure 3.36 – Protein alignments of AtTFL, AtFT, AtTSF with DtFT and DtTSF showing the key functional amino acids. ....	68
Figure 3.37 – Flowering time of T <sub>1</sub> generations of lines <i>AtCol0:DtFT</i> and <i>ft-10:DtFT</i> under LD. ....	69
Figure 3.38 – Flowering time data for <i>AtCol0:DtFT</i> T <sub>2</sub> generation under LD.....	70
Figure 3.39 – Flowering time data for <i>ft-10:DtFT</i> T <sub>2</sub> generation under LD.....	71
Figure 3.40 – Bolting data for <i>AtCol0:DtFT4</i> T <sub>3</sub> generation under SD.....	72
Figure 3.41 – Transgene amplification in complementation line <i>AtCol0:DtFT 4</i> .....	73
Figure 3.42 – Expression of <i>DtFT</i> transgene in <i>AtCol0:DtFT 4</i> T <sub>3</sub> generation under SD.....	73
Figure 3.43 – Flowering time data for <i>ft-10:DtFT 1</i> T <sub>3</sub> generation under SD.....	74
Figure 3.44 – Transgene amplification in complementation line <i>ft-10:DtFT1</i> .....	75
Figure 3.45 – Expression of <i>DtFT</i> transgene in <i>ft-10:DtFT 1</i> T <sub>3</sub> generation under SD. ....	75
Figure 3.46 – Flowering time data for <i>ft-10:DtFT 5</i> T <sub>3</sub> generation under SD.....	76
Figure 3.47 – Transgene amplification in complementation line <i>ft-10:DtFT 5</i> .....	77
Figure 3.48 – Expression of <i>DtFT</i> transgene in <i>ft-10:DtFT 5</i> T <sub>3</sub> generation under SD. ....	77
Figure 3.49 – Alignment report comparing <i>DtTSF</i> cds isolated from <i>Dt</i> with TSFfullLRcols used in complementation studies. ....	78
Figure 3.50 – Flowering time of T <sub>1</sub> generation of <i>DtTSF</i> transformed into <i>AtCol0</i> wild type and <i>ft-10</i> mutant. ....	79
Figure 3.51 – Inflorescence structures observed in <i>DtTSF</i> transformed plants.....	80
Figure 3.52 – Flowering time data by rosette leaf number at 1cm bolt for T <sub>2</sub> generation of DtTSF transformed lines.....	82
Figure 3.53 – Transgene amplification in complementation line <i>AtCol0:DtTSF 2</i> , <i>ft-10:DtTSF 7</i> and <i>ft-10:DtTSF 13</i> . ....	83
Figure 3.54 – Diurnal Time Course of <i>DtFT</i> . ....	83
Figure 3.55 – Diurnal Time Course of <i>DtTSF</i> . ....	84
Figure 3.56 – Expression of <i>DtFT</i> across development.....	85
Figure 3.57 – Expression of <i>DtTSF</i> across development.....	85
Figure 4.1 – Effect on bolting time of varying the length of 4°C cold treatment on <i>D. tenuifolia</i> seeds.....	101

Figure 4.2 – Effect of the different vernalization treatments at 4°C on expression of <i>DtFLC</i> in <i>D. tenuifolia</i> seed. ....	101
Figure 4.3 – Effect on bolting time of varying the length of 4°C cold treatment on <i>D. tenuifolia</i> four week old plants. ....	103
Figure 4.4 – Expression of <i>DtFLC</i> over development in vernalization treatments of different lengths. ....	103
Figure 4.5 – Effect on bolting time of varying lengths of cold treatment at 10°C on <i>D. tenuifolia</i> seeds.....	104
Figure 4.6 – Effect of the different vernalization treatments at 10°C on expression of <i>DtFLC</i> in <i>D. tenuifolia</i> seed. ....	105
Figure 4.7 – Effect on bolting time of varying the length of 10°C cold treatment on <i>D. tenuifolia</i> four week old plants. ....	106
Figure 4.8 – Effect on bolting time of varying photoperiod lengths in <i>D. tenuifolia</i> plants..	107
Figure 5.1 – MutMap method for fast SNP discovery in an EMS population.....	114
Figure 5.2 – Effect of EMS concentration on seed germination and number of days to bolt in Trial 1.....	116
Figure 5.3 – Effect of EMS concentration on seed germination and number of days to bolt in Trial 2.....	117
Figure 5.4 – Effect of EMS concentration on seed germination in EMS experiments 1 and 2. ....	118
Figure 5.5 – Bolting frequency of M <sub>2</sub> plants. ....	119
Figure 5.6 – Mean number of days to initiate bolting for EMS M <sub>3</sub> generation. ....	120
Figure 5.7 – Bolting frequency of M <sub>3</sub> plants. ....	121
Figure 5.8 – Bolting data for M <sub>2</sub> _BC <sub>1</sub> generation. ....	123
Figure 5.9 – Bolting data for M <sub>2</sub> _BC <sub>1</sub> _S <sub>1</sub> generation.....	125
Figure 5.10 – Normal and mottled, deformed leaf type.....	126
Figure 5.11 - Bolting data for M <sub>2</sub> _BC <sub>1</sub> _S <sub>1</sub> lines.....	127
Figure 5.12 – Segregating populations of M <sub>2</sub> _BC <sub>1</sub> _S <sub>1</sub> lines chosen for whole genome sequencing.....	129
Figure 6.1 – Diagram of a ZFN consisting of 2 sets of zinc fingers with the dimerised <i>FokI</i> endonuclease. ....	137
Figure 6.2 – A pair of TALENs showing orientation and function when aligned to a target gene. ....	138
Figure 6.3 – CRISPR/Cas9 system. ....	140
Figure 6.4 – Comparison of NHEJ (a+b) and HR (c+d) methods of DNA repair.....	141
Figure 6.5 – DtFT cds with sequences targeted by the CRISPR constructs highlighted.....	144
Figure 6.6 – Complete construction of AtU6-26-CRISPR-Cas9.....	146
Figure 6.7 – Alignment of CRISPR constructs showing insertion of CRISPR target regions. ....	153
Figure 6.8 – Flowering time of Arabidopsis plants transformed with CRISPR FT-96 construct.....	156
Figure 6.9 – Flowering time of Arabidopsis plants transformed with CRISPR FT-1 construct in experiments 1 and 2. ....	157

Figure 6.10 – Flowering phenotype in LD of late bolting Arabidopsis lines FT1_4 and FT1_5 compared to wild type (WT).....	158
Figure 6.11 – Flowering time of Arabidopsis plants transformed with CRISPR FT-2 construct in experiments 1 and 2. ....	159
Figure 6.12 – Flowering phenotypes of Arabidopsis transformed line FT2_3 compared to <i>AtCol0</i> wild type (WT). ....	160
Figure 6.13 – Flowering time of Arabidopsis plants transformed with CRISPR FT-136 construct.....	161
Figure 6.14 – Bolting phenotype of <i>D. tenuifolia</i> plants transformed with CRISPR FT96 and CRISPR FT2 constructs.....	162
Figure 6.15 – Alignment of CRISPR FT-1 transformed Arabidopsis <i>FT</i> gene sequences showing mutations. ....	163
Figure 6.16 – Alignment of CRISPR FT1 transformed and cloned Arabidopsis <i>FT</i> gene sequences showing mutations.....	164
Figure 6.17 – Transformed Arabidopsis protoplast fluorescing with GFP.....	169
Figure 6.18 – FACS result for Arabidopsis protoplasts transformed with GFP (isolated cells from 27/5/14). ....	169
Figure 6.19 – Transformed rocket protoplasts.....	170
Figure 6.20 – FACS result for transformed and non-transformed rocket protoplasts from protoplast isolation 27/05/14.....	170
Figure 6.21 – Rocket protoplasts at regeneration day 10.....	171
Figure 6.22 – Arabidopsis protoplasts at regeneration day 16.....	172
Figure 6.23 – Alignment of AtFT protein sequence against the translated FT1_4_5 clone sequence showing the additional amino acids. ....	173
Table 1.1– Five key flowering pathway genes in Arabidopsis chosen for investigation in rocket.....	15
Table 2.1 – Rocket seed received from gene bank IPK.....	20
Table 4.1 – Light regimes for photoperiod experiments 1 and 2.....	99
Table 6.1 – Quantities of protoplast cells, plasmid and PEG-calcium solution used for transformation of protoplast cells. ....	150
Table 6.2 – Media recipes needed for regenerating protoplasts .....	151
Table 6.3 – Number of transformed plants surviving hygromycin selection for each experiment.....	154
Table 6.4 – Summary of direct sequencing of FT PCR products from late bolting CRISPR transformed Arabidopsis plants .....	163
Table 6.5 – Sequencing results from cloning PCR products from late flowering CRISPR FT-1 transformed plants.....	165
Table 6.6 – Summary of protoplast isolations from Arabidopsis and rocket leaf material. ..	167
Table 6.7 – Blast results for 12 nucleotide sequence from PAM site for all CRISPRs to investigate possible off-target sites.....	174

## **Acknowledgements**

I would like to thank my supervisors Dr Stephen Jackson and Ms Sue Kennedy for their expert help and advice throughout the project and for willingness to accept me onto the PhD programme. To Dr Andrea Massiah and Professor Yiguo Hong who have also been supervisors for part of the project lifetime, I also give my sincere thanks for the support and encouragement they have given me. I have really appreciated the involvement of Dr Rosemary Collier and Dr Graham Teakle as my advisory panel members; they have both provided ideas and valuable insights on the project from a wider perspective, so thank you. For all the friends and colleagues I have had at Warwick HRI and Gibbet hill campus, especially the Thomas/Jackson lab group, those I have shared labs and offices with, and the support staff in hort. services, PBF and stores, I couldn't have made it through without your support, hard work and willingness to help when I have needed it, a huge thanks to you all.

To Stephanie Rippon, Alice Murphy, Anushree Choudhary and Meiling He who have worked with me as project students, thanks for your hard work and enabling me to achieve more in the time frame.

I thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of the Sequencing data and the University of Warwick's Bioinformatician Siva Samavedam for his work in making sense of all the data.

Huge thanks go to Elsoms Seeds Ltd as my industrial sponsors and to all the staff who I have worked alongside during the project, especially Richard Tudor, Adrian Dunford and Marie-Laure Bayard.

I would also like to thank Professor Hong and his research group at Hangzhou Normal University, particularly Dr Zhiming Yu, for the invitation to work with them on the CRISPR/Cas9 constructs. My time in the lab has been valuable to the project and the time spent at the university is something I will never forget.

Finally I would like to thank my family for their constant love and support during the PhD, it has been really tough at times and you have always been there for me. And to my Father God, without whom none of this would have been possible.

## **Declaration**

The work referred to in this thesis is my own, unless otherwise stated and has not been submitted for a degree at another university.

## Summary

Wild rocket (*Diplotaxis tenuifolia*) has recently become a popular salad leaf in the UK due to its peppery taste. It is grown widely in Italy but is now being produced in Spring and Summer in England. It is part of the *Brassicaceae* family and thus has a high level of homology at the DNA level to other *Brassica* species and *Arabidopsis*. This project aims to produce late bolting genotypes of rocket to incorporate into commercial breeding programmes. Delayed bolting is important as current varieties flower at unpredictable times and often earlier than desired. This is a problem because when rocket flowers it becomes unsaleable. Ethyl methanesulphonate (EMS) was used to generate a mutant population of rocket. Late bolting lines were selected and whole genome sequencing was used to identify single nucleotide polymorphisms (SNPs) induced by the EMS which may be a cause of the late bolting phenotype. Six flowering pathway genes have been isolated from rocket and have been tested to see if they can functionally complement *A. thaliana* knockout lines in these genes. Further work was carried out to investigate how these genes were expressed over diurnal and developmental time courses to understand their function in the flowering time pathway in rocket. Together, these results show that most of the flowering pathway genes isolated from rocket are functional orthologues of those in *Arabidopsis*. The photoperiodic and vernalization requirements of rocket were investigated and it was found that rocket does not have a vernalization requirement and is a facultative long day plant. Targeted mutagenesis using clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated system (cas9) was employed for the introduction of mutations into the *FLOWERING LOCUS T (FT)* gene in rocket and *Arabidopsis*. Many late flowering *Arabidopsis* lines were identified and sequencing revealed the successful manipulation of the *FT* gene. Two late bolting rocket lines were also identified. Overall, the project aims were achieved as late bolting rocket lines have been produced and will be incorporated into a rocket breeding programme.

## Abbreviations

<i>At</i> – <i>Arabidopsis thaliana</i>	LB – late bolting
<i>AtCol0</i> – <i>Arabidopsis thaliana</i> <i>Columbia</i> ecotype	LD – Long day
<i>AtLer</i> – <i>Arabidopsis thaliana</i> <i>Landsberg erecta</i> ecotype	M – mutagenised
BC – backcross	MS – Murashige and Skoog Basal Salt Mixture
CDS – coding sequence	NHEJ – non-homologous end joining
CO – CONSTANS	NLS – nuclear localisation signal
CRISPR/Cas9 – clustered regularly interspaced short palindromic repeat/CRISPR associated system	PAM – protospacer-adjacent motif
DSB – double strand break	PVX – potato virus X
<i>Dt</i> – <i>Diplotaxis tenuifolia</i>	QTL – quantitative trait loci
EMS – Ethyl Methanesulphonate	S – selfed
FACS – fluorescent activated cell sorting	SD – Short day
FDA – Fluorescein Diacetate	sgRNA – single guide RNA
FLC – FLOWERING LOCUS C	SOC1 – SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
FT – FLOWERING LOCUS T	SSD – single seed descent
gDNA – genomic deoxyribonucleic acid	TALENs – transcription activator-like effector nucleases
GFP – green fluorescent protein	TSF – TWIN SISTER OF FT
GI – GIGANTEA	TuMV – turnip mosaic virus
GM – genetic modification	UTR – untranslated region
GMOs – genetically modified organisms	UV – ultra-violet light
HR – homologous recombination	WT – wild type
InDel – Insertion or deletion	YFP – yellow fluorescent protein
	ZFNs – zinc finger nucleases
	ZT – zeitgeber time

## Chapter 1. Introduction

### 1.1. Project background

Global food security is a critical issue in the 21<sup>st</sup> century with a predicted population reaching nine billion by 2050. Recent estimates suggest that food production will need to double from current levels in order to sustain such a large number of people (2011). This huge increase in productivity will involve overcoming issues such as land shortages and water scarcity, fuel efficiency and climate change, with as little damage to the environment as possible to ensure a sustainable future. A key part of this will be the improvement of current crops, and the introduction of new crops, as well as reducing wastage and unnecessary crop losses. It is important not only to produce enough food for the growing population, but to provide a balanced diet consisting of the major food groups. Fruit and vegetables contribute a large number of essential vitamins and minerals and campaigns such as the UK Government ‘5-a-day’ initiative aim to encourage people to consume a least five portions of fruit and vegetables a day to maintain a healthy diet (<http://www.nhs.uk/livewell/5aday/pages/5adayhome.aspx>). Salad vegetables such as lettuce, spinach and rocket, particularly when bagged, can provide a convenient way of reaching the 5-a-day target, as they don’t need cooking and are often already washed. Wild rocket (*Diplotaxis tenuifolia*), despite its long history of use in the Mediterranean, is a relative newcomer to the UK and commercial cultivation. It therefore has a large scope for improvement. A major issue with commercial production of rocket is wastage and crop losses due to bolting as it renders the crop unsaleable. If delaying bolting, or increasing holding ability, can be achieved it would give growers more time to harvest the crop without a reduction in leaf quality as well as increasing sustainability through reducing wastage. This project aiming to delay bolting in rocket to improve quality and sustainability therefore has commercial benefits and was awarded a BBSRC CASE award in partnership with Elsoms Seed Ltd.

## 1.2. Overall aim

The aim of this PhD is to produce late bolting rocket lines, where the late bolting phenotype is robust despite differing environmental conditions, based on the hypothesis that EMS will introduce variation in bolting time into the population so that lines that are delayed in bolting time can be selected and characterised. Work will also be carried out to understand which genes are involved in the control of flowering time in rocket and isolate them to provide candidate genes which, if manipulated, could lead to a late bolting phenotype.

It is expected that the project will produce robust late bolting lines which can then be used to cross into a commercial line where leaf shape, size and taste are uniform across the crop, to create a potential market leading variety

## 1.3. Objectives

- Isolate homologues of flowering pathway genes
- Investigate the vernalization and photoperiodic flowering requirements of rocket
- Create an EMS population of rocket to select for late bolting lines
- Use high throughput genome sequencing to try to determine the causative SNPs within late lines
- Manipulate rocket flowering pathway genes using CRISPR technology

## 1.4. Introduction to rocket

### 1.4.1. Overview

Wild rocket (*D. tenuifolia*) has increased in popularity over the last 20 years in the leafy salads market (Bennett *et al.*, 2007; Chun *et al.*, 2013). In the UK alone, over 80 tonnes of rocket is consumed per week (Gill, 2008) and sourced from Italy, Portugal, Spain, UK and USA ([www.sainsburys.co.uk](http://www.sainsburys.co.uk)) is readily available year round in British supermarkets. It is known for its dark green, lobed leaves (Figure 1.1) which have a pungent peppery flavour (Mohamedien, 1994). Several species such as *D. tenuifolia* and *D. muralis* are eaten, usually in salads, whereas others such as *D. eruroides* are classed as weed species (Grillo *et al.*, 2012). Rocket has bright

yellow flowers which are abundant in pollen and emit an oilseed rape type aroma. Seed pods when produced have a double layer of seed which gives rise to the name of the genus (Pignone and Martinez-Laborde, 2011). *Diplotaxis* species are generally annual plants which have two seasons of growth, winter to spring and summer to autumn (Gresta *et al.*, 2010), but *D. tenuifolia* is a perennial (Pignone and Martinez-Laborde, 2011). Across the genus, *Diplotaxis* is a natural outcrosser (Eschmann-Grupe *et al.*, 2004), however one species, *D. viminea*, has been found to be self-fertilising (Grillo *et al.*, 2012)



Figure 1.1 – A *Diplotaxis tenuifolia* plant showing typical dark green and lobed leaves

#### 1.4.2. Phylogeny

Rocket, also known as roquette, arugula or rucola is a crop consisting of two genera, *Eruca* and *Diplotaxis*. Generally, *Eruca sativa* (salad rocket) is the popular rocket species in the USA and *D. tenuifolia* is known as wild rocket is more popular in Europe, however, due to the range of leaf shape and size in each, often the terms are used interchangeably. The genus *Diplotaxis* is found in the *Brassicaceae* family in the *Oleracea* clade (Arias and Pires, 2012) and is therefore closely related to species *Brassica rapa*, *B. juncea*, *B. napus*, *B. oleracea* as well as *Arabidopsis thaliana* (*Arabidopsis*) (Figure 1.2). This means that many genes and pathways that have been characterised in *Arabidopsis* and *Brassica* are likely to be common to rocket, so that there is a lot of relevant knowledge and sequence resource available for researchers aiming to improve *Eruca* and *Diplotaxis* crops. The similarity between *Brassica* and *Diplotaxis* is so close that lines have been introgressed to try and

introduce beneficial traits such as disease resistance into cultivated *Brassica* lines (Garg *et al.*, 2010; Gill, 2008). *Diplotaxis* is known to have 32 species (Warwick and Al-Shehbaz, 2006) which were originally found in Europe, Africa and Asia but several species have also spread to the Americas and Australia. Chromosome numbers vary across the genera as do phenotypes (Warwick and Al-Shehbaz, 2006). Phylogenetic trees of *Diplotaxis* have been constructed using chromosome numbers and cytodemes (Harberd, 1976), chloroplast-DNA restriction sites (Warwick *et al.*, 1992), Random Amplified Polymorphic DNA (RAPD) markers (Eschmann-Grupe *et al.*, 2004) and seed morphology (Grillo *et al.*, 2012).

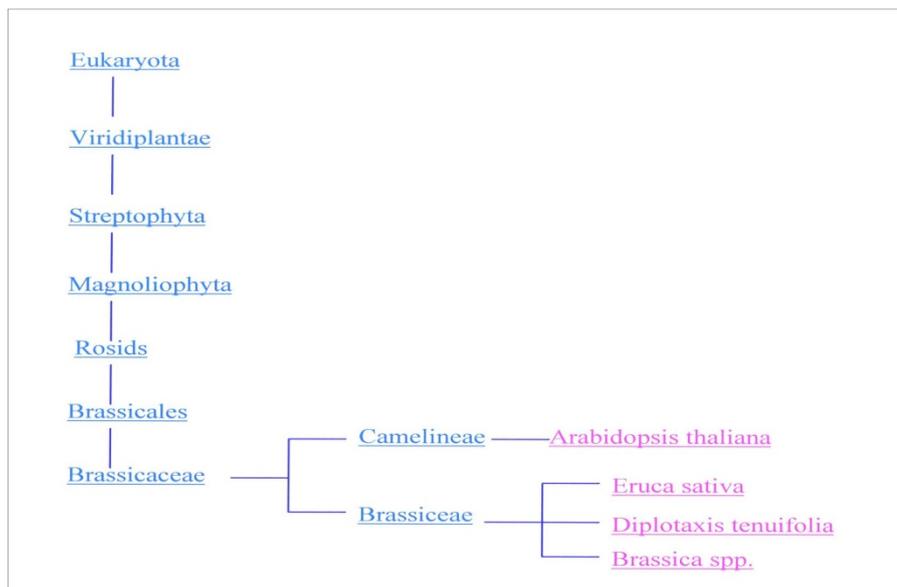


Figure 1.2 – Phylogenetic tree of rocket (Sayers *et al.*, 2011)

### 1.4.3. Cultivation

*D. tenuifolia* is cultivated in many regions of the world including the south of the UK (Weightman *et al.*, 2012). It is planted in lawns of up to 100 plants/m<sup>2</sup> as direct sowing at a high density of plants can give high yields, and it is harvested at about 2 cm above ground level using a lawnmower style machine (Figure 1.3) (Bianco and Boari, 1996; Esiyok, 1996). Depending on the climatic conditions and cost of production, rocket is grown both in the field and in glasshouses and usually drilled directly as seed. In ideal conditions, the crop can be ready for harvest within 25-35 days (Bell *et al.*, 2015). Rocket when cultivated, particularly in Mediterranean climates, can be harvested several times per crop, but research has shown that more than two cuts becomes economically unviable (Bianco and Boari, 1996) and can lead

to a quicker onset of flowering (Pimpini and Enzo, 1996). Nitrogen fertiliser at up to 100 kg/ha can be applied to aid growth of rocket especially for producing subsequent harvests (Baggio and Pimpini, 1995), but this must be balanced with the consideration that rocket readily takes up nitrates and nitrate levels can easily go above the recommended level for human consumption (Weightman *et al.*, 2012).



**Figure 1.3 – Lawnmower style harvesting machine used for harvesting leafy salad crops including rocket.**  
Photograph taken at Blackdown Growers, Radford Semele, Warwickshire, UK.

#### **1.4.4. Taste and health**

The characteristic peppery taste of rocket is due to the glucosinolate content, which also makes it a good candidate for pharmaceutical applications. *D. tenuifolia* has a history of being used in traditional medicine for various ills such as digestive disorders and diabetes (Yaniv, 1994). It is usual for the leaves of rocket to be eaten, however research by Falleh *et al.* (2013) has shown that the flowers contain the highest levels of beneficial chemicals with antioxidant and antibacterial activities. Bell *et al.* (2015) found that Dimeric-4-mercaptobutyl and Glucosativins are the glucosinolates in highest abundance in *D. tenuifolia* at the time of harvest. Hall *et al.* (2015) investigated the glucosinolate changes in pre- and post-harvesting and discovered that the level and type of glucosinolates changes from the first to second harvests. Glucosinolates accumulate in the vacuoles of the cells to keep them

spatially separated from the enzyme myrosinase which is in the cytosol. When tissue is broken down, the myrosinase reacts with the glucosinolates producing the volatile isothiosinates (Bennett *et al.*, 2006).

#### **1.4.5. Breeding**

Research into rocket and its breeding programmes became centralised and more focussed with the creation of the rocket genetic resources network in the 1990s (Pignone and Martinez-Laborde, 2011). Most research to date has involved analysing glucosinolate content, flavour and nutrient content, as well as post-harvest degradation and packaging improvements (Bell *et al.*, 2015; Bennett *et al.*, 2007; Lokke *et al.*, 2012; Weightman *et al.*, 2012). There is therefore an opportunity to investigate some less understood aspects of the crop, such as flowering behaviour and underlying genetic pathways controlling growth and development.

#### **1.4.6. Bolting**

When conditions are favourable, a network of genes triggers the change from vegetative growth to the reproductive phase. Visually this is first identified by the formation of floral buds followed by the upward growth of the stem, known as bolting. The buds then mature and the flowers open.

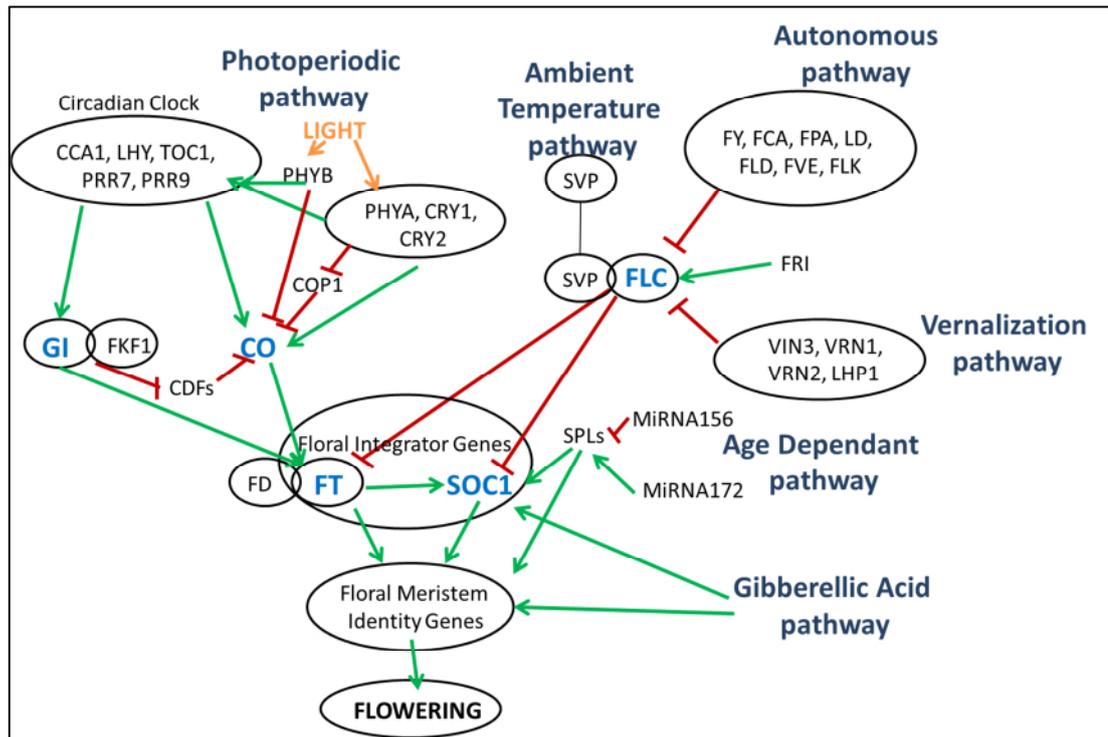
Bolting/flowering is a problem in many leafy vegetables as it can reduce the quality of the product and the potential yield (Abou-Elwafa *et al.*, 2011; Bluemel *et al.*, 2015). For example, in lettuce, the production of floral buds is linked to a change in secondary metabolites, such as sesquiterpene lactones causing an unpleasant taste (Sessa *et al.*, 2000). Although this isn't the case in rocket, bolting is still undesirable due to the industry standards in marketable rocket which doesn't include floral parts. The whole crop therefore has to be discarded if individuals begin to bolt (Raw Material Specification, Bakkavor Group Ltd).

A delay in bolting of current rocket varieties by seven days or more would be commercially beneficial as crop scheduling could be improved and less wastage would occur, with the consumer still being satisfied with the product quality and uniformity (S. Kennedy, Elsoms Seeds Ltd, personal communication).

## 1.5. Introduction to flowering

### 1.5.1. Overview

Regulation of flowering time involves a complex system of genes and networks which allows a plant to respond to internal and external signals to control the timing of the transition from the vegetative phase to the reproductive phase (Srikanth and Schmid, 2011). The timing of this transition is crucial to enable reproduction to occur when the plant is at its fittest and the environmental conditions are most favourable (Thomas *et al.*, 2006). Most research into flowering time has been conducted in the model plant *Arabidopsis*, through the generation and testing of many gene knockout mutant lines and transgenic overexpression lines to elucidate gene function (Corbesier and Coupland, 2005). Currently there are six key pathways that regulate flowering time which are: the photoperiodic, autonomous, vernalization, gibberellic acid, age dependent and ambient temperature pathways (Figure 1.4) (Brambilla and Fornara, 2013; Fornara *et al.*, 2010; Jarillo and Pineiro, 2011; Srikanth and Schmid, 2011). All of these pathways converge at the floral pathway integrator genes *FLOWERING LOCUS T (FT)* and *SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1)* which activate the floral meristem identity genes. When flowering is induced the shoot apical meristem (SAM) changes from forming vegetative tissues such as leaves to form flowers.



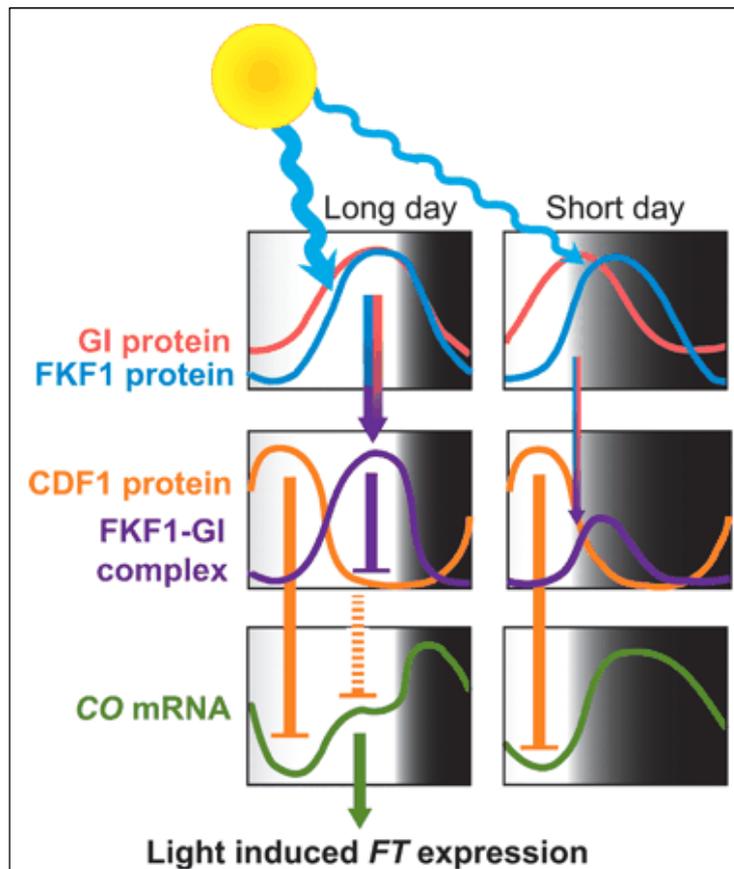
**Figure 1.4 – Flowering time regulatory pathways in Arabidopsis.**

Green arrows represent gene activation, red T-bars represent gene repression, genes in blue are those analysed in this study. Gene abbreviations: *CCA1* - *CIRCADIAN CLOCK ASSOCIATED 1*, *LHY* - *LATE ELONGATED HYPOCOTYL*, *TOC1* - *TIMING OF CAB EXPRESSION 1*, *PRR7&9* - *PSEUDO-RESPONSE REGULATORS*, *GI* - *GIGANTEA*, *FKF1* - *FLAVIN-BINDING, KELCH REPEAT, F-BOX 1*, *CDF* - *CYCLING DOF FACTOR*, *PHYA* - *PHYTOCHROME A*, *PHYB* - *PHYTOCHROME B*, *CRY1* - *CRYPTOCHROME 1*, *CRY2* - *CRYPTOCHROME 2*, *COP1* - *CONSTITUTIVELY PHOTOMORPHOGENIC*, *CO* - *CONSTANS*, *FD* - A bZIP transcription factor, *FT* - *FLOWERING LOCUS T*, *SVP* - *SHORT VEGETATIVE PHASE*, *FLC* - *FLOWERING LOCUS C*, *LD* - *LUMINIDEPENDENS*, *FLD* - *FLOWERING LOCUS D*, *FLK* - *FLOWERING LOCUS KH DOMAIN*, *FRI* - *FRIGIDA*, *VIN3* - *VERNALIZATION INSENSITIVE 3*, *VRN1* - *VERNALIZATION 1*, *VRN2* - *VERNALIZATION 2*, *LHP1* - *LIKE HETEROCHROMATIN PROTEIN 1*, *SPL* - *SQUAMOSA PROMOTER BINDING LIKE*, *SOC1* - *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*. (Adapted from Corbesier and Coupland (2005), Feng *et al.* (2011), Sawa and Kay (2011), Srikanth and Schmid (2011))

### 1.5.2. Photoperiodic pathway

The photoperiodic pathway results from the integration of the circadian clock, the endogenous 24 hour rhythm, and the day length (Suarez-Lopez *et al.*, 2001). Two of the key genes in this pathway are *CONSTANS (CO)* and *GIGANTEA (GI)* (An *et al.*, 2004; Suarez-Lopez *et al.*, 2001). In long days (LD), *GI* is expressed under the control of the circadian clock early on in the day and *GI* protein forms a complex with FLAVIN-BINDING, KELCH REPEAT, F-BOX (*FKF1*) which targets CYCLING DOF FACTORS (*CDFs*) for degradation. *CDF* expression is regulated by the circadian clock and is induced by the clock genes *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*, and *LATE ELONGATED HYPOCOTYL (LHY)* in the morning. *CDFs* repress the expression of *CO* by binding to its promoter (Song *et al.*,

2015). When CDFs are degraded, *CO* can then be expressed, towards the end of the LD and overnight (Figure 1.5) (Miller *et al.*, 2008; Mizoguchi *et al.*, 2005). The accumulation of CO protein at the end of a LD is stabilized by the action of the photoreceptors PHYTOCHROME A (PHYA) and CRYPTOCHROMES 1&2 (CRY1 & CRY2), and the high levels of CO trigger the expression of *FT* which promotes *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* and the floral integrator genes which induces flowering (Kim *et al.*, 2008). In short days (SD) in *Arabidopsis*, flowering is delayed. This is due to the timing of *GI* expression not being synchronised with *FKF1* with the result that the CDF isn't degraded and *CO* expression is much lower (Figure 1.5) (Ito *et al.*, 2012), so *FT* expression isn't activated (Sawa and Kay, 2011). Further discussion of the photoperiodic pathway can be found in Chapter 4, and the key photoperiodic genes *CO*, *GI* and *FT* are discussed in more depth including their sequence, structure and function in rocket in Chapter 3.



**Figure 1.5 – Role of key photoperiod genes *GI*, *FKF1*, *CDF* and *CO* in *Arabidopsis* under long day and short day conditions.**

The differential expression of genes involved with floral transition by photoperiodic response culminating on *FT* activation. Figure from Sawa *et al.* (2007)

### 1.5.3. Autonomous pathway

The autonomous pathway is made up of seven major genes. These are *LUMINIDEPENDENS* (*LD*), *FLOWERING LOCUS D* (*FLD*), *FLOWERING LOCUS KH DOMAIN* (*FLK*), *FLOWERING LOCUS CA* (*FCA*), *FLOWERING LOCUS PA* (*FPA*), *FLOWERING LOCUS Y* (*FY*) and *FLOWERING LOCUS VE* (*FVE*). These are all able to promote flowering by repressing *FLOWERING LOCUS C* (*FLC*) which is a floral repressor that represses *FT* and *SOC1* (Simpson, 2004). *LD* contains a homeodomain-like domain and is involved in the regulation of *FLC* through histone modification (Domagalska *et al.*, 2007) and interaction with SUPPRESSOR OF FRI 4 (*SUF4*) (Kim *et al.*, 2006a). *FLD* encodes a histone demethylase protein, and regulates *FLC* by chromatin remodelling (He *et al.*, 2003). *FLK* encodes an RNA-binding protein with three K-homology-type RNA-binding domains and functions to repress *FLC* through RNA processing, although this is yet to be fully elucidated (Cheng *et al.*, 2003; Lim *et al.*, 2004). *FCA* is an RNA recognition motif (RRM) and WW domain containing RNA binding protein (Macknight *et al.*, 1997). It acts to downregulate *FLC* by binding with *FY* or *FPA* and *FLD* (Liu *et al.*, 2007). *FCA* has also been shown to interact with *FVE* (a homolog of putative retinoblastoma-associated proteins in animals) through histone deacetylation in the ambient temperature pathway (Blazquez *et al.*, 2003). *FPA* contains three RRM and as well as functioning with *FCA*, has a role in RNA polyadenylation (Hornyik *et al.*, 2010). *FY* is a 3' end-processing factor interacting with *FCA* for *FLC* regulation and also in the autoregulation of *FCA* expression (Simpson *et al.*, 2003). Normally during vegetative growth the levels of *FLC* are very high which prevents flowering. The autonomous pathway genes work to repress the expression of *FLC* so levels decrease. When this happens, *FLC*'s repression of the downstream genes *FT* and *SOC1* is lifted and floral induction is then possible (Amasino, 2010).

### 1.5.4. Vernalization pathway

Vernalization is defined as "the acquisition or acceleration of the ability to flower by a chilling treatment" (Chouard, 1960). It is normally a plant's response to the cold winter months which enable it to flower when the climate becomes more favourable and flowering occurs more quickly when subjected to the cold period. In the model plant *A. thaliana*, the vernalization requirement varies in different ecotypes. Some

ecotypes are summer annuals which have lost their vernalization requirement and have no need for cold exposure to flower. Others which are winter annuals will flower very late if they are not vernalized, but with the vernalization treatment, they are able to flower much earlier (Alonso-Blanco and Koornneef, 2000; Michaels and Amasino, 2000). The expression of two genes, *FLC* and *FRIGIDA (FRI)*, determine whether a plant is able to respond to vernalization or not. As mentioned above *FLC* is a repressor of flowering, and *FRI* regulates *FLC* expression. In plants where *FRI* is expressed, expression levels of *FLC* are high, so increases the time it takes for flowering to occur. This is due to *FLC* blocking transcription of *FT* and *SOC1* by binding to CArG boxes on their chromatin (Helliwell *et al.*, 2006). When *FRI* is expressed at a low level, or is absent, *FLC* expression is not expressed at a high level and flowering is earlier as it is not repressed (Johanson *et al.*, 2000). Cold treatment causes levels of *FLC* expression to reduce in the plant through epigenetic changes in the *FLC* gene which affect its expression. Key histones in *FLC* are methylated in response to vernalization which prevent transcription and cause the reduction in *FLC* levels (Michaels, 2009). When the plant is returned to a higher temperature, the change is stable enough to enable floral promotion rather than a return to high *FLC* levels (Boss *et al.*, 2004). Studies using *Arabidopsis* mutants *vernalization1 (vrn1)* and *vrn2* showed that these genes were responsible for maintaining the low level of *FLC* during, and after, vernalization. These mutants, when returned to higher temperatures following vernalization, did not flower as expected as the level of *FLC* had risen once more (Gendall *et al.*, 2001; Levy *et al.*, 2002). This memory of cold treatment can be maintained during tissue culture as it is mitotically stable (Wellensiek, 1964). However, the epigenetic changes are not transferred through the germ line, so the subsequent generation are once again very late or unable to flower without going through a period of cold treatment (Jarillo and Pineiro, 2011). Experiments investigating the vernalization response in rocket are described in Chapter 4 and the structure and function of rocket *FLC* in Chapter 3.

#### **1.5.5. Age dependant pathway**

The age dependent pathway results in an increased ability of a plant to respond to floral induction signals as it ages. The SQUAMOSA PROMOTER BINDING LIKE (SPL) transcription factors are repressed by *microRNA (miR) 156* early in development but its expression decreases as the plant matures, allowing an increase

in the levels of the SPLs. These then promote the expression of *SOC1* and the floral meristem identity genes *LEAFY (LFY)* and *APETALA 1 (API)* (Cong *et al.*, 2013a; Jarillo and Pineiro, 2011). *miR172* promotes the expression of SPLs and it increases in abundance as the plant ages. *miR172* also induces *FT* through the repression of *FT* repressors *TEMPRANILLO1 (TEM1)*, *TEM2*, *SCHLAFMÜTZE (SMZ)*, *SCHNARCHZAPFEN (SNZ)*, *TARGET OF EAT1 (TOE1)*, *TOE2* and *TOE3* (Castillejo and Pelaz, 2008; Mathieu *et al.*, 2009). It has been suggested that starch metabolism has a role to play in the age dependent pathway through trehalose-6-phosphate. This acts as a signal of sucrose levels in the plant and regulates its response through *miR156* and SPLs (Lunn *et al.*, 2006; Matsoukas *et al.*, 2012; Wahl *et al.*, 2013).

#### **1.5.6. Ambient temperature pathway**

The ambient temperature pathway involves the response of *SHORT VEGETATIVE PHASE (SVP)* to the environmental temperature. SVP forms a complex with FLC to ensure the repression of *FT* and *SOC1* continues thus preventing flowering when temperatures are low (Fornara *et al.*, 2010; Jarillo and Pineiro, 2011). Flowering occurs more rapidly when the temperature is higher as this complex no longer forms. Studies have shown that Arabidopsis *svp* mutants are able to flower early regardless of temperature as *FT* is not repressed by SVP (Fornara *et al.*, 2010). Studies on the involvement of *miR156* and *miR172* in the ambient temperature pathway have linked these to the differing regulation by SVP (Lee *et al.*, 2010). The autonomous pathway genes *FCA* and *FVE* have also been implicated in temperature sensitive regulation of flowering (Srikanth and Schmid, 2011), and work by Kim *et al.* (2012) has also suggested that *miR156* acts through *SPL3* to regulate *FT* expression in response to temperature.

#### **1.5.7. Gibberellic acid pathway**

The gibberellic acid pathway induces flowering through the action of bioactive gibberellic acid (GA). A high level of GA promotes the expression of *SOC1* and *LFY*, leading to floral induction. This was discovered through a study of rice which was infected with the fungus *Gibberilla fujikuroi*. The fungus synthesises gibberellins, which causes a fast growth rate and lodging in the infected rice plant (Phinney, 1983). Subsequent research found that Arabidopsis mutants which impair

the production of GA are delayed in flowering, especially when grown in SD conditions (Fornara *et al.*, 2010). This phenotype was able to be reversed when GAs are added exogenously (Hedden and Phillips, 2000). GA perception in plants is through the GIBBERELLIC INSENSITIVE DWARF 1 (GID1) receptor and high GAs induce the expression of *FT* in long days (Griffiths *et al.*, 2006). This is through the GA mediated degradation of DELLA proteins which delay flowering by reducing the expression of *FT* and *TSF* (Galvao *et al.*, 2012). GAs have also been implicated in the activation of the other floral integrator genes *SOC1* and *LFY* (Blazquez *et al.*, 1998; Moon *et al.*, 2003).

#### **1.5.8. Role of flowering pathway genes in crop species**

The manipulation of flowering pathways has been key in crop domestication, allowing the spread of species from high temperature, intermediate photoperiod regions to cold temperature, seasonal regions where the day length changes from short to long days depending on the time of year (Nakamichi, 2015). With the work in *Arabidopsis* on flowering pathways, the understanding of the genes involved in the control of flowering in different crop species is growing. The photoperiod pathway genes *CO* and *FT* and their role together in promoting flowering, has been found to be conserved across monocots and dicots (Jung and Mueller, 2009). The rice *CO* homolog *Heading date1* (*Hd1*) has been shown to promote flowering through its *FT* homolog *Heading date3a* (*Hd3a*), even though rice is a SD plant (Izawa *et al.*, 2002). Rice also has *GI* homolog *OsGI* which is regulated by the circadian clock and induces the expression of *Hd1* (Hayama *et al.*, 2003). Potato also has homologs of *CO* and *FT* which induce tuberization as well as flowering (Rodriguez-Falcon *et al.*, 2006). Barley *Photoperiod- H1* (*Ppd-H1*) is a homolog of the *Arabidopsis* circadian clock gene *PSEUDO-RESPONSE REGULATOR* (*PPR7*) and variation of this gene has given the species the plasticity to grow under different photoperiodic conditions (Turner *et al.*, 2005). Variation in key vernalization pathway genes has also been important in the development of crop species. In *Brassicaceae*, homologs of *FLC* have been found, and account for the majority of spring sown and winter sown varieties in their response to vernalization (Jung and Mueller, 2009). The only non-*Brassicaceae* species where *FLC* homologs have been discovered are root chicory (*ciFL1*) (Perilleux *et al.*, 2013), sugar beet (*BvFL1*) (Reeves *et al.*, 2007) (although this is being debated (Vogt *et al.*, 2014)), and lettuce (*LsFLC1-5*) (A. Massiah,

personal communication). Vernalization responses in other species are also due to variation at the *VRN* locus, such as the non-vernalizing Canadian spring wheat through *Vrn-A1a* (Cockram et al., 2007; Iqbal et al., 2007). The continued application of knowledge gained from Arabidopsis flowering time studies into crop species will further our understanding of the molecular mechanisms and provide a wealth of possibilities for crop breeding today and in the future.

## **1.6. Project strategy**

The work in this project falls into two main categories. The first is to understand the flowering pathway in rocket, so transferring knowledge gained from Arabidopsis into this crop species. The second is to delay the onset of flowering in rocket to produce late-bolting lines for incorporation into commercial breeding programmes. The results in Chapters 3-6 will expand on each of these points and explain the progress made in this project.

### **1.6.1. Understanding the flowering pathway in rocket**

Much research has been done to explore the underlying mechanisms involved in the floral transition in the model species Arabidopsis. A complex network of interlocking pathways and genes has been found, and new players are being discovered all the time. However, it is important that this knowledge gets applied in crop systems to enable higher yielding, better quality and delayed bolting varieties to be developed. Five key genes in the Arabidopsis flowering pathway were chosen for investigation in rocket (Table 1.1).

**Table 1.1– Five key flowering pathway genes in Arabidopsis chosen for investigation in rocket.**

<b>Gene</b>	<b>Function</b>	<b>Location in flowering pathway</b>	<b>Locus in Arabidopsis</b>
<i>CO</i>	Regulation of flowering under long days	Photoperiodic pathway	AT5G15840
<i>GI</i>	Regulates <i>CO</i> by complexing with FKF1	Photoperiodic pathway	AT1G22770
<i>FLC</i>	Repressor of flowering	Autonomous pathway, vernalization pathway, ambient temperature pathway	AT5G10140
<i>SOCI</i>	Floral induction	Floral integrator	AT2G45660
<i>FT</i>	Floral induction	Floral integrator	AT1G65480

BLAST searches of these Arabidopsis genes revealed homologues in *Brassica* species. Sequence information from the Arabidopsis and *Brassica* sequences for each gene were used to isolate the gene homologs in rocket. Chapter 3 describes the results of the flowering time gene isolation in rocket following the specific aims:

- i) Isolation of the five key flowering pathway genes in rocket
- ii) Use the isolated coding sequences (CDS) of the genes for functional complementation studies in Arabidopsis
- iii) Investigate the diurnal and developmental expression pattern of each gene in rocket

In order to develop our understanding of the control of flowering in rocket, experiments were designed to determine the vernalization and photoperiodic requirements for flowering which are not currently known. Chapter 4 describes the

results for both the vernalization and photoperiod experiments in rocket following the specific aims:

- i) Establish whether rocket has a vernalization requirement through subjecting plants to two different vernalization temperatures for different lengths of time and monitor the expression of *FLC*.
- ii) Explore the flowering response of rocket in different length photoperiods

### **1.6.2. Use of mutagenesis to delay the onset of flowering in rocket**

The use of various mutagenesis methods is wide spread, both in plant research and in crop breeding. It enables new traits to be discovered by changing the activity of genes within the plant. Ethyl methanesulphonate (EMS) has been used in *Arabidopsis* to provide mutant plants which have enabled the understanding of the function of many genes. The use of EMS in crop species such as tomato (Gady *et al.*, 2009) and rice (Abe *et al.*, 2012) has enabled breeders to select desirable traits for improving crop varieties. Despite EMS introducing mutations into a plants genome, crop varieties bred using this technique are deemed non-GM. EMS has also been used successfully on rocket previously (Kenigsbuch *et al.*, 2014), so for these reasons it was the method of choice for this project. This work aimed to produce an EMS-mutagenesied population for rocket which could be screened for late bolting phenotypes. Whole genome sequencing will be employed to allow single nucleotide polymorphisms (SNPs) to be identified in order to try to link one of them to the late bolting phenotype observed. Chapter 5 describes the results for the EMS treatment and late bolting trait selection as well as SNP identification following the specific aims:

- i) Create an EMS population of rocket and select for late bolting phenotypes
- ii) Use backcrossing to reduce the load of non-causative background SNPs
- iii) Identify SNPs which differ between wild type and selected late bolting lines

EMS is a random method of introducing mutations into a plants genome. Technology has advanced over the last 20 years to provide methods of targeted mutagenesis (Bortesi and Fischer, 2015). This has given the ability to choose specifically which genes are mutated. This has the potential of speeding up breeding programmes as multiple backcrosses to the wild type to reduce the number of

background mutations causing pleiotropic phenotypes are not needed. However, the sequences of the target genes in crop species are needed before mutagenesis can take place but with the cost of high-throughput sequencing reducing all the time, this is becoming more achievable. In this project, with the knowledge gained through the isolation of key flowering genes described in Chapter 3, the *FT* gene was chosen as an initial candidate for targeted mutagenesis. A collaboration was set up with Professor Yiguo Hong and his research group at Hangzhou Normal University, China who had the knowledge and experience of using clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated system (CRISPR/Cas9) for targeted mutagenesis. This was therefore our method of choice. Chapter 6 outlines results of using this system to mutate *FT* in both Arabidopsis and rocket following the specific aims:

- i) Design the CRISPR target regions to *FT* and build the CRISPR/Cas9 construct (carried out on an exchange visit to Hangzhou Normal University)
- ii) Transformation of CRISPR/Cas9 constructs into Arabidopsis and rocket
- iii) Trial and optimise protoplast isolation and regeneration as a potential method of transient transformation of CRISPR/Cas9 constructs in Arabidopsis and rocket
- iv) Identify late bolting transformed lines and sequence *FT* gene for evidence of targeted mutagenesis

## **1.7. Elsoms Seeds Ltd and CASE award**

This project is a BBSRC CASE award funded PhD in partnership with Elsoms Seeds Ltd. Elsoms Seeds Ltd was founded in 1844 and is a small family-owned company. It has a key role in seed production and sales in the UK and for the last 30 years has been operating a seed treatment business. They have a strong history in plant breeding, which is growing and encompasses partnerships with academia and other industrial companies to encourage research and innovation into breeding new crop varieties. Their rocket breeding programme began in 2009 using both *Eruca sativa* and *Diplotaxis tenuifolia* leading to the collaboration with the University of Warwick and the development of this project. This CASE award partnership is aiming to build on the research of flowering time in crop plants in the Jackson Lab at the University

of Warwick, produce late bolting rocket lines for Elsoms Seeds Ltd to incorporate into their rocket breeding programme and provide the student Jemma Taylor with expertise as a plant scientist and experience of working in industry.

## Chapter 2. Materials and Methods

### 2.1. Materials

#### 2.1.1. Arabidopsis

*Arabidopsis thaliana* Columbia-0 (Col-0) and *ft-10* (Col-0 background) mutant seed was provided by Dr Aaron Abbott (Warwick University), *A. thaliana* Landsberg erecta (Ler) seed from Mrs Alison Jackson (Warwick University), FRI+ *flec-3 null* (Col-0 background) mutant seed from Dr Andrea Massiah (Warwick University), *soc1-2* (Col-0 background) mutant seed from Dr Richard Immink (Wageningen University, Netherlands) and *gi-3* (Col-0 background) and *co-2* (Ler background) mutant seed directly from the Nottingham *Arabidopsis* Stock Centre (NASC) (<http://arabidopsis.info/>), where all the above seed was originally sourced.

#### 2.1.2. Rocket (*Diplotaxis tenuifolia*)

##### 2.1.2.1. Single Seed Descent

Elsoms Seeds Ltd provided seed which originated from a commercial *Diplotaxis tenuifolia* variety 'Voyager' (Tozer Seeds Ltd), but had been bred through three generations of single seed descent for uniformity of leaf shape and colour, and for late bolting. The seed received was single seed descent 3 (SSD3). Fifty SSD3 seeds were sown at University of Warwick, Wellesbourne campus (latitude 52°12') to produce large numbers of SSD4 seed for the EMS mutagenesis (section 2.1.6.3). Bolting data from the SSD3 plants was recorded and 2000 SSD4 seeds were collected from six SSD3 plants which bolted around the average number of days for the population. This was used to generate the EMS population. Seed from a single SSD3 plant that had an average bolting time was sown to create the SSD4 population, and bolting dates recorded. This was repeated for SSD5 and SSD6. Bolting data collected for SSD6 showed a smaller range of bolting time. SSD6 seed was threshed and used in subsequent experiments.

##### 2.1.2.2. Gene Bank Seed

*D. tenuifolia* seed was also sourced from a gene bank (IPK, Gaterslaben, Germany) (Table 2.1) for the EMS trial experiments as the SSD4 seed was not available at the time of the experiment.

**Table 2.1 – Rocket seed received from gene bank IPK.**

Accession	Species	Origin
DIPLO3	<i>Diplotaxis tenuifolia</i>	Botanic Garden, Siena, Italy
DIPLO4	<i>Diplotaxis tenuifolia</i>	Botanic Garden, Liege, Belgium

### **2.1.3. Scoring plants for bolting**

When a plant changes from its vegetative state to its reproductive phase, the shoot apical meristem transforms from producing leaves, to producing floral parts. When observed closely, small flower buds can be seen, indicating that the phase change has occurred (Figure 2.1A). For this work, when this is seen the plants is scored for initiation, and the date recorded. The speed of bolting, or the number of days to initiate is worked out using the initiation date and subtracted the date of sowing or the date of germination. It is an important measurement as it is the closest observation that can be made to the change in gene regulation causing the reproductive phase to start. The second measurement taken from the plant is bolting taken as when there is growth of the shoot apical meristem to above the level of the foliage; usually about 10 cm (Figure 2.1B). This is a very obvious sign that the change to the reproductive phase has occurred, and is the measurement used by growers and industry to tell whether a crop has bolted and therefore unsaleable. The date is recorded when the shoot has reached 10 cm, and the number of days to bolt worked out using the date of sowing or germination date for that plant. Scoring for initiation is more labour intensive than scoring for bolting, but taking both allows a comparison across experiments and to records from growers and industry about how various rocket lines are performing against one another. The third way of scoring for bolting is by counting the number of leaves as used by Kenigsbuch *et al.* (2009; 2014). This is like the method of scoring bolting in *Arabidopsis* as it assumes the plant rosette stops producing leaves when the reproductive phase begins. However, this method is very labour intensive for rocket which produces many more leaves before bolting, so has only been used once in this work.



**Figure 2.1 – Scoring rocket plants for bolting.**

(A) Arrow points to the small floral buds observed and scored in an initiated plant. (B) Arrow points to large floral buds at the top of a 10 cm bolt when scoring is recorded for a bolted plant.

#### **2.1.4. Rocket pollination techniques**

##### **2.1.4.1. Insect pollination**

Where the plant population was large and grown in the field and selfed seed was needed, flies were used to aid pollination (Figure 2.2). Individual plants were bagged using perforated plastic bread bags and sealed at the bottom using plant ties. House Fly (*Musca domestica*) pupae were inserted into the bag and allowed to hatch. The flies then pollinated any open flowers present in the bag during their lifetime; seed from these flowers would be selfed seed and was collected to propagate that line.



**Figure 2.2 – Bagged plant with flies to aid pollination.**  
Yellow arrow points to a fly.

#### **2.1.4.2. Bud pollination**

To aid pollination where the plant population was small and a specific cross was desired, bud pollination was used. An inflorescence stalk was selected and any open flowers and seed pods were removed (Figure 2.3A). Using fine-nosed forceps, the small central buds were removed and the apex destroyed. The top half of the sepals and petals were removed and the stamens taken out revealing the stigma from the biggest 4-6 buds (Figure 2.3B). An open flower with a high level of mature pollen from either the same plant where self-seed is required, or from a wild type plant where backcross seed is required was brushed against the stigma of all exposed buds (Figure 2.3C). The inflorescence was bagged to prevent any unwanted pollen transfer (Figure 2.3D). The silique emerges after one week if the bud pollination is successful and develops to a full seed pod after four weeks.



**Figure 2.3 – Bud pollination of *Diplotaxis tenuifolia*.**

A) Single inflorescence where open flowers will be removed, B) Cutting buds to remove sepals, petals and stamen, C) Brushing pollen onto exposed stigmas, D) Bagging bud pollinated inflorescence

### **2.1.5. Seed threshing**

*D. tenuifolia* seed heads were collected in perforated bread bags after drying on the plant for four weeks. These bags were stored in dry conditions until threshing. To thresh the seed from the seed heads, bags were emptied carefully into a 1.7 mm sieve and moved until the plant material broke down. Large size waste was discarded and small chaff and seed was then sieved through a 1 mm sieve twice. Remaining plant debris of a smaller size than the seed was sieved out using a 425  $\mu\text{m}$  sieve before the remaining seed collected into a paper bag. Seed was then stored in dry conditions at 2-10°C.

*A. thaliana* plants were dried in paperbags and threshed in a similar way to *D. tenuifolia* seed but using 2x 500  $\mu\text{m}$  sieve. Seed was collected into 1.5ml Eppendorf tubes and stored in dry 2-10°C conditions.

### **2.1.6. Rocket ethyl methanesulphonate (EMS) mutagenesis**

#### **2.1.6.1. EMS treatment**

*D. tenuifolia* seed was counted into aliquots of 200 seeds. These were sealed inside muslin pockets to allow the EMS to penetrate without seed being lost. These were placed into a 500 ml Duran bottle containing 24 ml 10% Tween20 and 216 ml sterile H<sub>2</sub>O. The bottle was placed into a shaking incubator for 15 minutes at 25°C, 150 rpm. The seed pockets were rinsed four times with H<sub>2</sub>O before filling the bottle with 300 ml sterile H<sub>2</sub>O and put on the shaking incubator for 5 minutes at 25°C, 150 rpm.

This was repeated three times. Each seed pocket was then placed into a 250 ml Duran bottle containing sterile H<sub>2</sub>O and EMS at the appropriate concentration. One batch of 200 seeds did not receive any EMS to allow a control comparison of experimental effects. Each bottle was put at 25°C for 16 hours on a shaking platform at 150 rpm. The EMS solution was poured off and inactivated with 0.1 M NaOH/20% STS for 24 hours before pouring down the fume cupboard sink. The bottles were filled with 150 ml H<sub>2</sub>O, swirled and emptied six times before refilling with 150 ml H<sub>2</sub>O and placing in the incubator for 15 minutes at 25°C, 150 rpm. The H<sub>2</sub>O was poured off and a final continual wash of H<sub>2</sub>O was done for three hours to remove any further traces of EMS. Muslin bags were opened and seeds were sown at one per cell in 98 cell trays of wet peat and covered with vermiculite before germinating at 25°C, 16 hour photoperiod for one week.

#### **2.1.6.2. EMS trial experiments**

Two trials of EMS treatment were done to optimise the concentration of EMS to use for rocket. *D. tenuifolia* gene bank seed (IPK, Gaterslaben, Germany) was used because the SSD4 seed was not yet available (section 2.1.1). The same EMS protocol was used throughout as described in section 2.1.6.1.

For Trial 1, gene bank seeds DIPLO 3 (Table 2.1) were used and the following concentrations trialled:

EMS concentration
0 (control)
10 mM (~0.1%)
20 mM (~0.2%)
30 mM (~0.3%)

For Trial 2, gene bank seeds DIPLO 4 (Table 2.1) were used and the following concentrations trialled:

EMS concentration
0 (control)
5 mM (~0.05%)
10 mM (~0.1%)
20 mM (~0.2%)

Germination tests were performed at 25°C, 16 hour photoperiod for seven days. The number of germinated seed was then recorded. Plants were transplanted at three weeks old into 5 litre pots of Levington M2 soil and grown to maturity at Elsoms Seeds Ltd, Spalding (Latitude 52°79') in a 16 hour photoperiod glasshouse at 22°C. The results of these trials enabled a decision to be made on what concentration of EMS to use in Experiments 1 and 2.

#### **2.1.6.3. EMS experiments**

Ten batches of 200 SSD4 seeds (section 2.1.1) were treated at a concentration of 10 mM EMS on 16/05/12 using the method described in section 2.1.6.1. Germination data was recorded as for the trial experiments. Trays were then transferred to 16 hour photoperiod at 20°C conditions for further growth. At six weeks old (03/07/12), these were put into a cold frame to acclimatise to outside conditions and at eight weeks old (17/07/12), were planted into the field (section 2.1.7.2).

A second experiment was performed on another 2000 SSD4 seeds (sections 2.1.1&2.1.6.1) with an EMS concentration at 20 mM on 27/06/12. As previously, germination data was recorded and plants transferred to 16 hour photoperiod at 20°C conditions for further growth. At four weeks old (25/07/12), these were put into a cold frame to acclimatise to outside conditions and at six weeks old (09/08/12), were planted into the field (section 2.1.7.2).

#### **2.1.6.4. Production of EMS M<sub>2</sub> generation**

Harvested seeds from 1127 plants across both M<sub>1</sub> experiments were threshed. Of these, only 332 yielded M<sub>2</sub> seed. Seed was also threshed from 10 untreated control lines. From each of these lines, up to six seeds were sown onto wet peat and germinated at 20°C, 16 hour photoperiod. Germination was checked after seven

days and any lines which had not germinated were re-sown. After 14 days, only 223 lines had one or more seed germinate. Three weeks after sowing, plants were put into a cold frame to acclimatise to outside conditions for one week before planting into Spanish polytunnels. Plants which were later bolting than the wild types were dug up and potted into 5 litre pots and put into the glasshouse at 20°C, 16 hour photoperiod.

#### **2.1.6.5. Production of EMS generations M<sub>2</sub>\_BC<sub>1</sub> and M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub>**

Individual flowers on late bolting M<sub>2</sub> plants were backcrossed to wild type plants of SSD6, the purest plant line to date (section 2.1.1), using bud pollination (section 2.1.4.2). Seed was set and the plants allowed to dry out before collecting and threshing the seed (section 2.1.5). Threshed seed was sown into F2+S soil in p40 cells and covered with vermiculite. Trays were placed into a glasshouse compartment at University of Warwick Phytobiology facility at 22°C with 16 hour photoperiod. These were covered with propagator lids until one week after germination. Seedlings were transplanted into 5 inch pots of M2 soil at four weeks from sowing. Initiation of bolting and 10 cm bolt was scored for every plant. Flowers were bud pollinated and bagged to ensure selfed seed was generated.

#### **2.1.6.6. Production of M<sub>3</sub> generation**

Selfed seed was produced for the M<sub>3</sub> generation so late bolting M<sub>2</sub> plants were bagged to prevent pollen transfer from different plants. M<sub>3</sub> seed was threshed before up to eight seed were sown onto wet peat. Seeds were germinated at 20°C, 16 hour photoperiod and grown for three weeks. These were put into a cold frame to acclimatise for one week before transplanting into the field (section 2.1.7.2).

### **2.1.7. Growth conditions**

#### **2.1.7.1. Glasshouse trials**

Rocket plants in the Single Seed Descent experiments were grown at the University of Warwick, Wellesbourne campus at a set temperature of 20°C with 16 hours light, 8 hours dark where high pressure sodium lamps (Philips 400 W SON-T bulbs) provided supplementary lighting when light levels dipped below 1500  $\mu\text{mol}/\text{m}^2/\text{s}$ . Plants were watered regularly and nutrient feed (Vitax 2:1:4 NPK) added weekly once plants had reached eight true leaves.

Rocket plants in the EMS experiment M<sub>2</sub>\_BC<sub>1</sub> and M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub> were grown at the University of Warwick Phytobiology Facility at a set temperature of 22°C with 16 hours light, 8 hours dark where high pressure sodium lamps (Philips 400 W SON-T bulbs) provided supplementary lighting when light levels dipped below 1500 μmol/m<sup>2</sup>/s. Plants were watered regularly and nutrient feed (Vitax 2:1:4 NPK applied at 0.5 g/l) added weekly once plants had reached eight true leaves.

### **2.1.7.2. Field trials**

Rocket plants in the EMS population were grown at University of Warwick, Wellesbourne campus (latitude 52°12'). Seed was sown onto peat blocks, covered with vermiculite and watered. M<sub>1</sub> seed was germinated for one week at 25°C with 16 hours light, 8 hours dark in a Procema controlled environment cabinet before transferring to glasshouse conditions as described in 2.1.6.1. M<sub>2</sub> and M<sub>3</sub> seeds were sown in the same way but germinated in a seedling raising glasshouse under natural conditions with no temperature or lighting supplementation. At four weeks from sowing, seedlings were transplanted into the field. M<sub>1</sub> plants were planted with Imidiclopid capsules for insect pest control. M<sub>2</sub> and M<sub>3</sub> plants were treated initially with Intercept 70WG for pest control when transplanted. Two weeks later treatment of Steward at a rate of 250 g/ha was used at weekly intervals until flowering where treatment was biweekly, to control prevalence of flea beetle and thrips. These populations were also laid out using an alpha block randomisation method (Patterson *et al.*, 1978). The M<sub>2</sub> population was planted into Spanish polytunnels to increase the growing temperature and prevent waterlogging which was an issue when growing the M<sub>1</sub> population. The plants were watered regularly using drip tape irrigation. M<sub>1</sub> and M<sub>3</sub> populations were also watered when necessary using overhead sprinklers. Insect pollination was used to ensure self-seed was produced in M<sub>1</sub> generation (section 2.1.4.1). For the M<sub>2</sub> generation, selected late bolting lines were dug up and transferred to the glasshouse where bud pollination for self and backcross seed was used (section 2.1.4.2).

### **2.1.7.3. Growth of plants in controlled environment cabinets**

Controlled environment cabinets (Versatile plant growth MLR-352, Panasonic Co. Ltd) were used for rocket 10°C vernalization, diurnal and photoperiod experiments and Arabidopsis complementation T<sub>2</sub> and T<sub>3</sub> experiments. Cabinets were set to 22°C constant temperature with 16 hour light and 8 hour dark conditions for long days and

22°C constant temperature with 8 hour light and 16 hour dark conditions for short days, except in vernalization and photoperiod experiments. Fluorescent bulbs provide light at an average of 124  $\mu\text{mol}/\text{m}^2/\text{s}$ .

## **2.2. Standard laboratory methods**

### **2.2.1. DNA extraction**

Rocket genomic DNA was extracted using CTAB extraction method adapted from Stewart and Via (1993) on frozen leaf material, which was ground to a powder using a Dremel Drill with a 1.5 ml Eppendorf tube drill bit. 300  $\mu\text{l}$  CTAB B buffer (100mM Tris/Cl pH 8.0, 1.4M NaCl, 20mM EDTA, 2% hexadecyltrimethyl ammoniumbromide) was added and homogenised. Samples were incubated at 65°C for 30 minutes and centrifuged. The supernatant was removed into a new tube and twice extracted using 300  $\mu\text{l}$  Chloroform:Isoamyl alcohol (24:1) and centrifugation. The aqueous top phase was transferred each time. 300  $\mu\text{l}$  CTAB C buffer (1% hexadecyltrimethyl ammoniumbromide, 10mM EDTA, 50mM Tris/Cl pH 8.0) was added and left overnight at room temperature. On day two, the tubes were centrifuged and the pellet dissolved in 400  $\mu\text{l}$  1 M CsCl. DNA is then precipitated using 100% Ethanol and centrifuged. The pellet was then washed twice using 70% Ethanol before drying and resuspending in 25  $\mu\text{l}$  TE buffer (pH 8) with RNase A (Invitrogen) (20  $\mu\text{g}/\text{ml}$ ). When small amounts of leaf material were used, the CTAB method was reduced to half volume, and the final pellet resuspended in 15  $\mu\text{l}$  TE buffer (pH 8) with RNase A (Invitrogen) (20  $\mu\text{g}/\text{ml}$ ).

### **2.2.2. RNA extraction and cDNA synthesis**

Rocket RNA extraction was done using the Z6 extraction buffer (containing 8 M guanidine hydrochloride) method (Logemann *et al.*, 1987). Leaf material was ground to a powder using liquid nitrogen and a pestle and mortar. The powder was transferred to a 1.5 ml Eppendorf tube and Z6 buffer plus 2-mecaptoethanol (50mM final concentration) added. This was homogenised using a Dremel Drill with a 1.5 ml Eppendorf tube drill bit before the rest of the published method was followed. 5  $\mu\text{g}$  RNA was DNase treated using Ambion® Turbo DNA-free™ DNase and then resuspended in 12  $\mu\text{l}$  DEPC treated H<sub>2</sub>O. 1  $\mu\text{g}$  of DNase treated RNA was

synthesised into cDNA using an Invitrogen ThermoScript cDNA synthesis kit or BioRad iScript cDNA synthesis kit.

### **2.2.3. Primer design**

Sequences for each of the *CO*, *FLC*, *FT*, *GI* and *SOCI* genes from *A. thaliana*, *Brassica napus*, *B. oleracea* and *B. rapa* were aligned using DNA Lasergene 9 core suite MegAlign package. Primer3Plus software (available online <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) was used to design degenerate primers for each gene and manufactured by Invitrogen. Once the gene sequences had been isolated for each gene, specific primers were then designed and manufactured by Sigma Aldrich (Appendix A).

### **2.2.4. Polymerase Chain Reaction (PCR)**

PCR was performed using gDNA and cDNA, using the primers at 0.6  $\mu$ M final concentration for degenerate primers and 0.3  $\mu$ M final concentrations for specific primers. Novagen® Kod Hot Start PCR kit was used unless otherwise stated. The standard reaction volume was 20  $\mu$ l and PCR conditions were 95°C for 2 minutes and then the following cycled 35 times: 95°C for 20 seconds, annealing temperature for the primer pair for 10 seconds, 70°C for the extension time of 15 seconds/kb. A final temperature of 70°C is held for 4 minutes. For products larger than 3 kb, a Novagen® Kod Xtreme™ PCR kit was used with the standard reaction volume of 25  $\mu$ l. PCR conditions were 94°C for 2 minutes and then the following cycled 35 times: 98°C for 10 seconds, annealing temperature for the primer pair for 30 seconds, 68°C for the extension time of 1 minute/kb. A final temperature of 68°C was held for 4 minutes.

### **2.2.5. PCR product visualisation and purification**

Products (3  $\mu$ l) were run out on a 1% (w/v) agarose (AGTC BioProducts)/1xTAE gel stained with Ethidium bromide or GelRed (Biotium) at 100 V for 20 minutes. Gels were visualised on a UV lightbox (Syngene G-box with GeneSnap software) and single band PCR products were purified using a Qiagen PCR purification kit and finally eluted in 20  $\mu$ l H<sub>2</sub>O. To isolate bands from reactions with multiple PCR products, 20  $\mu$ l was run on a 1% (w/v) agarose/1xTAE gel for products larger than 1.2 kb and for those smaller than 1.2 kb, a 2% (w/v) agarose/1xTAE gel was used.

The correct band was excised on a UV light box and purified using a Qiagen gel extraction kit and eluted in 20 µl H<sub>2</sub>O.

#### **2.2.6. DNA sequencing**

1.2 µl DNA was quantified using a NanoDrop®ND-1000 Spectrophotometer (Thermo Scientific). DNA was sequenced using Big-Dye and a 3130xl Genetic Analyzer (Applied Biosystems) DNA Sequencing machine, or couriered to GATC Biotech (Germany) where reactions were performed and sequence analysed on ABI 3730xl machines. M13F and M13R primers were used to sequence cloned products, and PCR primers used to sequence PCR products (Appendix A). Results were then compared to *Arabidopsis* and *Brassica* species using NCBI BLAST programme (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and contigs generated using DNA Lasergene 9 core suite SeqMan package.

#### **2.2.7. GeneRacer™**

Primers were designed to the putative gene sequences in rocket using parameters defined in the Invitrogen GeneRacer™ Kit (RNA ligase-mediated rapid amplification of 5' and 3' cDNA ends (RLM-RACE)). 5' and 3' Expressed Sequence Tag (EST) ends were obtained using a GeneRacer™ Kit (RLM-RACE) for each gene, cloned into pGEM-T easy and sequenced as described previously.

#### **2.2.8. Molecular cloning**

##### **2.2.8.1. Blunt end cloning**

For products up to 3 kb, pGEM-T easy vector system (Promega) was used for cloning. Before products were purified, A-tailing was performed by adding 1 µl of 10 mM dNTPs and 1 µl 500U Taq polymerase (Invitrogen) and heating to 70°C for 30 minutes to add an Adenine nucleotide. This enables efficient ligation due to the pGEM-T vector having a Thymine overhang. Purification and quantification using a NanoDrop®ND-1000 Spectrophotometer followed before ligating into pGEM-T easy vector at a 3:1 ratio using T4 DNA ligase overnight at 4°C. 2 µl of the ligation product was added to 20 µl *Escherichia coli* (*E. coli*) EC100 cells and electroporated. 1 ml of SOC buffer was added and the cell culture put at 37°C for 1 hour. The cell culture was plated onto LB (10:5:10)/ 1.5% (w/v) agar plates with 100 µg/ml ampicillin, 20 µg/ml isopropylthio-β-D-galactoside (IPTG) and 20 µg/ml X-GAL. The plates were incubated at 37°C overnight. White colonies which contain inserts

were picked and added to 40 µl H<sub>2</sub>O. This re-suspension was used as the template for the following PCR using an Invitrogen Taq polymerase PCR kit. The primers were M13F x M13R which anneal to the vector sequence either side of the insert (Appendix A.2.1.2). The conditions were 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 52°C for 30 seconds and 72°C extension of 20 seconds/kb of the expected product size. A final cycle of 72°C for 10 minutes was performed.

For products larger than 3kb, the product was ligated into a pCR®-BluntII-TOPO® vector using a Zeroblunt® TOPO® PCR cloning kit (Invitrogen). This was done by adding 1 µl purified PCR product with 1 µl of pCR®-BluntII-TOPO® vector mix with 3.7 µl salt solution (Zeroblunt® TOPO® PCR cloning kit) and made up to a final volume of 20 µl with H<sub>2</sub>O. The reaction was put at room temperature overnight. 20 µl H<sub>2</sub>O was added and 1 µl of the total ligation product was added to One Shot® TOP10 Electrocomp™ *E. coli* electrocompetent cells and electroporated. 250 µl of SOC buffer was added and the cell culture put at 37°C for 1 hour. This was then plated onto LB (10:5:10)/1.5% (w/v) agar plus 50 µg/ml kanamycin and incubated at 37°C for 16-18 hours. Colonies were picked and screened for the correct insert using the same method outlined above.

#### **2.2.8.2. Plasmid DNA extraction**

Colonies containing the correct insert were used to inoculate 5 ml LB broth with the appropriate antibiotic at the correct final concentration and grown overnight at 37°C. A Plasmid DNA Extraction kit (Thermo Scientific) was used to extract the vector and insert DNA from the *E. coli* cells following the manufacturer's instructions. The elution was done using 50 µl EB buffer and the DNA concentration was quantified before sequencing. When extracting plasmid DNA from *Agrobacterium tumefaciens* cells, a 10 ml LB broth was used to grow the culture overnight at 28°C. Centrifugation was done at 3500 rpm for 15 mins before pouring off the supernatant. The pellet was resuspended using P1 buffer from a Qiagen QIAprep Spin Miniprep kit with the addition of lysozyme at 4 mg/ml. This was then incubated at 37°C before following the rest of the manufacturer's instructions and eluting in 50 µl EB buffer at 65°C.

#### **2.2.8.3. GATEWAY cloning**

Products for cloning were generated using primers with *att* sites specific to the Gateway® cloning system, and amplified using PCR from pGEM-T easy or TOPO

cloned products. Each was quantified using a NanoDrop®ND-1000 Spectrophotometer and diluted to 150 ng/μl. Gateway® BP Clonase® II Enzyme mix (Invitrogen) (2 μl) was mixed with 1 μl 150 ng/μl PCR product (containing *att* sites), 1 μl 150 ng/μl pDONR207 vector (donated by Dr Karl Morris, University of Warwick) and 6 μl TE buffer. The reaction was incubated at 25°C for 16-18 hours. Ligation product was added to *E. coli* EC100 electrocompetent cells (Cambio Ltd) and electroporated. LB (10:5:10)/1.5% agar plus 15.6 μg/ml Gentamycin plates were used to select for insert-containing colonies. A PCR screen of selected colonies was performed using the gene specific primers, followed by plasmid DNA extraction from positive colonies and sequencing of plasmids using the same primers. The plasmid DNA from the colonies containing the correct sequence was used with the Gateway® LR Clonase® Enzyme mix (Invitrogen). The enzyme (2 μl) was added to 1 μl 150 ng/μl plasmid DNA from the correct sequence carrying vector from the BP reaction and 1 μl 150 ng/μl pB2GW7 vector (Karimi *et al.*, 2002)(donated by Mrs Rachel Clewes, University of Warwick) and 6 μl TE buffer. The reaction was incubated at 25°C for 16-18 hours. The product (1 μl) was added to 20 μl *E. coli* EC100 electrocompetent cells (Cambio Ltd) and electroporated. LB (10:5:10)/1.5% (w/v) agar plus 100 μg/ml Spectinomycin plates were used to recover insert containing colonies. A PCR screen of selected colonies was performed followed by plasmid DNA extraction and sequencing of positive clones.

### **2.2.9. Functional complementation of Arabidopsis mutants**

pB2GW7 plasmid DNA containing cDNAs representing the CDS of each rocket gene was transformed into *A. tumefaciens* strain c58pGV3101 using electroporation. 500 μl SOC buffer was added and the cell culture incubated at 28°C for 2-3 hours. This was then plated onto LB (10:5:10)/1.5% (w/v) agar plus 15.6 μg/ml Gentamycin, 100 μg/ml Spectinomycin and 12.5 μg/ml Rifampicin. The plates were incubated at 28°C for 48 hours. The clones were checked for the correct plasmid insertion by inoculating 10 ml LB (10:5:10) plus 15.6 μg/ml Gentamycin, 100 μg/ml Spectinomycin and 12.5 μg/ml Rifampicin and incubating for 16-18 hours at 28°C before extracting the DNA using a Plasmid DNA Extraction kit (Thermo Scientific) (section 2.2.8.2). The DNA was sequenced. When positive clones had been found, 500 ml LB (10:5:5) plus 25 μg/ml Gentamycin, 100 μg/ml Spectinomycin and 12.5 μg/ml Rifampicin was inoculated using 5 ml cell culture of *A. tumefaciens* strain

c58pGV3101 transformed with the plasmid. This was incubated at 28°C for 16-18 hours. The culture was centrifuged and the supernatant removed. 500 ml 5% (w/v) sucrose solution was used to resuspend the cells and 100 µl silwet L-77 added before dipping the inflorescences of the selected plants (Clough and Bent, 1998). The plants were sealed in a bag for 24 hours before putting at 22°C with 16 hours light, 8 hours dark in the Phytobiology facility (section 2.1.7.1). T<sub>1</sub> seed was harvested and sown onto *Arabidopsis* mix soil (Levington F2s : sand : vermiculite fine grade 6:1:1). BASTA (Ammonium glyfosinate (150 g/L)) soil soaking was used at 1:1000 as the selection method. The first treatment was given and the trays were covered and placed at 4°C for three days. These were removed and put under a propagator lid in a 16 hour photoperiod at 22°C. Four further treatments were done before transplanting. Transformed plants were transplanted into p24 pots containing *Arabidopsis* mix soil and leaf number and number of days to bolt were recorded when the primary bolt was 1 cm. Leaf material was collected for subsequent experiments. T<sub>2</sub> seed was collected from each plant and threshed. A screen of bolting time was done on this generation by sowing 50 T<sub>2</sub> seed from functionally complemented T<sub>1</sub> lines onto *Arabidopsis* mix soil. These were stratified at 4°C for three days in the dark and grown in Panasonic controlled environment cabinets as described above. Leaf number and number of days to bolt were recorded when the primary bolt was 1 cm and data analysed for 3:1 ratio of early/late bolting time:wild type bolting time. Leaf material was collected for subsequent experiments.

#### **2.2.10. Real time PCR gene expression analysis**

Primers suitable for real time PCR were designed to each flowering time gene isolated from rocket. This was done using Primer3Plus software (available online at <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>). Three housekeeping genes, *TIP41*, *α-tubulin* and *CLATHERIN ASSOCIATED COAT SUBUNIT (CACCS)*, were also identified and isolated from rocket using degenerate primers and PCR. Real Time primers were then designed to the isolated rocket sequence for each housekeeping gene. cDNA was generated from 1 µg DNase treated RNA from leaf material harvested weekly over the course of the experiment. cDNA standards were made from a mixture of 1 µl of RNA from 3 time points in each treatment or from PCR products for each gene used in the experiments isolated from cDNA. Standards were then used to create a tenfold dilution series from 1 to

$10^{-3}$ . 0.5  $\mu$ l cDNA was used in the reaction mix along with 5  $\mu$ l iTaq<sup>TM</sup>Universal SYBR<sup>®</sup> Green Supermix (Bio-Rad Laboratories Ltd., UK). The primers were optimised for each gene by testing at final concentrations of 0.1  $\mu$ M, 0.2  $\mu$ M and 0.4  $\mu$ M. Each was added to the reaction mix and cDNA and made up to a total of 10  $\mu$ l with DEPC H<sub>2</sub>O. 9  $\mu$ l was added to each well in a 384 well plate and loaded into the CFX384 Touch<sup>TM</sup> Real-time PCR machine (Bio-Rad Laboratories Ltd., UK). Bio-Rad CFX manager 3.0 software was used to run the programme. The protocol used was 95°C for 3 minutes followed by 44 cycles of 95°C for 10 seconds and 59°C for 30 seconds and a final step of 95°C for 10 seconds. The melt curve analysis then followed where the machine ramped from 65°C to 95°C at 0.5°C increments of 5 seconds each. Using the Bio-Rad CFX manager 3.0 software, the results were assessed and then analysed using Biogazelle qBase Plus software version 2.5 (<http://www.biogazelle.com/qbaseplus>). The data from the samples was normalised to the TIP41,  $\alpha$ -tubulin and CACS housekeeping genes to get the normalised expression for each flowering time gene of interest over the experiment.

## Chapter 3. Flowering Time Gene Isolation

### 3.1. Introduction

This chapter describes the isolation and characterisation of five key genes in the flowering time pathway of rocket. These genes are the rocket homologues of the Arabidopsis *CO*, *GI*, *FLC*, *SOC1* and *FT* genes and were chosen because they are located at key convergent points in the flowering time pathway (Chapter 1), thus mutations in any of these genes would lead to a change in the flowering time phenotype of the plant. Identifying, or creating, mutations in these key flowering genes would be one way to achieve the aim of producing plants which are delayed in bolting.

Each gene was isolated from rocket cDNA and gDNA by PCR using primers designed to sequences in the Arabidopsis and *Brassica* genes. PCR and RNA ligase-mediated rapid amplification of 5' and 3' cDNA ends (RLM-RACE) (GeneRacer™ Kit, Invitrogen) were used to amplify gene fragments until the entire gene sequence was obtained. Using these cloned rocket genes, complementation studies of the respective Arabidopsis mutants were carried out to determine whether or not the rocket flowering candidate genes function to complement the mutant and are able to recover a normal flowering time phenotype. The complete CDS for each gene also enabled specific primers to be designed for use in real time PCR, therefore experiments were undertaken to investigate the expression of these key genes in diurnal and developmental time courses.

#### 3.1.1. CONSTANS

The *CONSTANS* (*CO*) gene is a key player in the photoperiodic pathway and is regulated by the circadian clock. The AtCO protein has two B-box domains at the N-terminus and a CCT domain at the C-terminus (Miller *et al.*, 2008) forming a nuclear zinc finger transcription factor (An *et al.*, 2004). It is part of a large gene family of 17 members in Arabidopsis which are grouped according to the number and sequence of the B-box domains (Robson *et al.*, 2001). Expression of *CO* oscillates in a 24 hour cycle and functions to promote flowering under long day conditions (Suarez-Lopez *et al.*, 2001). This occurs as a result of circadian regulation where degradation of CDFs by the GI/FKF1 complex allows *CO* to be

expressed. The CO protein is also regulated post-translationally being stabilised at the end of the day through the photoreceptor genes *PHYA*, *CRY1* and *CRY2* (Valverde *et al.*, 2004). A high level of CO at the end of the day promotes *FT* expression leading to flowering. *CO* has been found to be highly conserved across the plant kingdom, with homologues identified in both dicots such as Arabidopsis (Suarez-Lopez *et al.*, 2001), Brassica (Robert *et al.*, 1998) and soybean (Fan *et al.*, 2014), and in monocots such as maize (Miller *et al.*, 2008) and rice (Yano *et al.*, 2000).

### **3.1.2. GIGANTEA**

GIGANTEA (GI) is a large protein made up of 1173 amino acids in Arabidopsis. It has not been found to be related to any other genes, but is conserved across higher plants (Mishra and Panigrahi, 2015; Mizoguchi *et al.*, 2005). Studies by Huq, Tepperman, and Quail (2000) used a GFP fusion demonstrating that GI is located to the nucleus. In long days, GI forms a complex with FKF1 which promotes flowering by removing the repression of *CO* by CDF1 (Shrestha *et al.*, 2014). *CO* then activates *FT* leading to flowering. It has also been found that GI binds directly to the promoter of *FT* preventing *SVP*, *TEM1* and *TEM2* from repressing *FT* so flowering can occur (Sawa and Kay, 2011). Homologues of *AtGI* have been found in other crop species where the GI function appears to be conserved, such as rice (Hayama *et al.*, 2002; Hayama *et al.*, 2003) and members of the Brassica family (Xie *et al.*, 2015). It would be interesting to see if rocket also has a *GI* homologue and whether it plays a similar role in the photoperiodic pathway leading to flowering.

### **3.1.3. FLOWERING LOCUS C**

*FLOWERING LOCUS C (FLC)* is a member of the MADS transcription factor family and is a repressor of flowering. *FLC* lies at a key convergent point between the vernalization and autonomous pathways. Both pathways promote flowering by inhibiting the expression and accumulation of *FLC* thus releasing the inhibition on the floral integrator genes *FT* and *SOC1* (Michaels and Amasino, 2001). There are seven genes implicated in the inhibition of *FLC* in the autonomous pathway which are *FLK*, *FLD*, *FVE*, *FPA*, *LD*, *FY* and *FCA*. These genes function to modify histones, bind RNA or remodel chromatin of *FLC* itself or other genes which then affect *FLC* (Baeurle *et al.*, 2007). *FLC* also inhibits flowering in the ambient

temperature pathway by binding to *SVP*, another MADS box family gene (Hartmann *et al.*, 2000). This interaction represses the expression of *SOCI* and *FT* so flowering is prevented under conditions which are sub-optimal for reproduction (Li *et al.*, 2008).

Sequence variation in *FLC* genes in Brassica species (Kim *et al.*, 2007; Tadege *et al.*, 2001; Wu *et al.*, 2012) and Arabidopsis (Alonso-Blanco and Koornneef, 2000; Strange *et al.*, 2011) has led to a wide variety of ecotypes which differ in their ability to, and speed at which they, flower (Lin *et al.*, 2005). The vernalization requirement of rocket and the role of *FLC* are further explored in Chapter 4.

#### **3.1.4. SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1**

*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOCI)* is a member of the MADS transcription factor family (Immink *et al.*, 2012). Work by Lee *et al.* (2000) demonstrated that *SOCI* expression increases slowly over development until flowering occurs. Homologues of *SOCI* have been found across the plant kingdom, where the role in flowering promotion has been retained (Kim *et al.*, 2003; Rantanen *et al.*, 2014; Shen *et al.*, 2014). *SOCI* is activated by *FT* through the photoperiodic pathway, and also by the age dependent and gibberellic acid pathways and by *FLC* through the vernalization, ambient temperature and autonomous pathways (Balanz *et al.*, 2014; Samach *et al.*, 2000; Tao *et al.*, 2012). It is therefore a key central integrator in the flowering pathway. As well as being involved in flowering, *SOCI* also has a role in floral organ formation and preventing indeterminacy and secondary growth (Melzer *et al.*, 2008).

#### **3.1.5. FLOWERING LOCUS T**

The *FLOWERING LOCUS T (FT)* and *FT*-like genes belong to a family of Phosphatidylethanolamine-Binding Protein (PEBP) containing genes (Kardailsky *et al.*, 1999). Members of this family in Arabidopsis include *FLOWERING LOCUS T (FT)*, *TWIN SISTER OF FT (TSF)*, *BROTHER OF FT (BFT)*, *MOTHER OF FT (MFT)*, *TERMINAL FLOWER (TFL)* and *ARABIDOPSIS THALIANA CENTRORADIALIS HOMOLOGUES (ATC)* (Kim *et al.*, 2013). The *FT* protein was identified to be ‘florigen’ which is expressed in leaves and moves to the SAM with the aid of *FT-INTERACTING PROTEIN1 (FTIP1)* (Liu *et al.*, 2012). It forms a complex with *FD* (a bZIP transcription factor) at the SAM and promotes flowering

by activating the floral identity genes such as *SOCI*, *LFY* and *API* (Abe *et al.*, 2005). TSF is also an inducer of flowering but is weaker than FT and also binds to FD at the SAM to promote flowering. In Arabidopsis, TSF is the closest homologue of FT at 82% (Yamaguchi *et al.*, 2005). MFT was also found to be a weak an inducer of flowering (Mimida *et al.* 2001). TFL is a strong repressor of flowering and related genes BFT and ATC are also repressors but much weaker in effect than TFL (Mimida *et al.*, 2001; Yoo *et al.*, 2010). Unlike Arabidopsis which has one *FT* and one *TFL* gene, *B. napus*, another close relative of rocket, has six *FT* and four *TFL* paralogues. Each of these has a different effect on the regulation of flowering (Guo *et al.*, 2014; Wang *et al.*, 2012).

## **3.2. Methods**

### **3.2.1. Gene isolation**

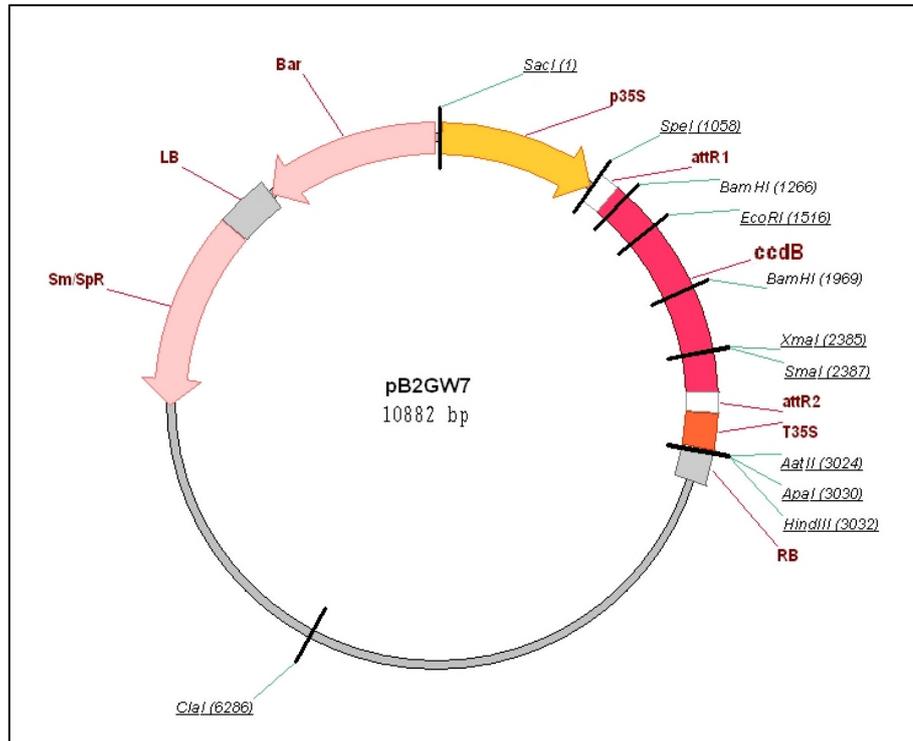
Leaf material was collected from *D. tenuifolia* (*Dt*) Single Seed Descent 4 (SSD4) plants at different time points during a long day and at different developmental stages when it was expected that the genes of interest would be highly expressed. RNA was extracted from all the leaf material individually using the Z6 buffer method (Logemann *et al.*, 1987) as outlined in section 2.2.2. DNase treatment was carried out on RNA from each sample using Ambion® Turbo DNA-free™ DNase. 1 µg DNase-treated RNA was synthesized into cDNA using a BioRad iScript cDNA synthesis kit. Degenerate primers were designed and used to amplify regions of *Dt* specific sequence for all genes of interest. PCR products were purified and sequenced as outlined in sections 2.2.5 and 2.2.6. Once these sequences had been obtained, new primers specific to each *Dt* gene (Appendix A.1) were designed for GeneRacer™ Kit (RLM-RACE) (Invitrogen, see section 2.2.7). Further primers were designed for each gene using the sequence data from GeneRacer™ results which covered the start and stop codons, and other primers designed to 5' and 3' untranslated regions (UTR) (START/STOP and 5'/3' primer pairs, see Appendix A.1) to enable the amplification by PCR of the entire coding sequence (CDS) for each gene. DNA was extracted from leaf material of *Dt* SSD4 plants and using the START/STOP and 5'/3'-UTR primer pairs, the full length genomic DNA (gDNA) sequence was PCR'd, isolated and sequenced for each gene of interest using Kod Hot Start and Kod Xtreme™ PCR kits (Novagen®) as described in sections 2.2.4-

2.2.6. Additional sequence needed to be isolated for the *FT*-like genes; this was done using a PolyATtract® mRNA Isolation System (Promega). RNA was extracted using the Z6 buffer method outlined above from leaf material collected from Dt SSD6 plants at ZT16 close to the bolting date when the *FT* gene is likely to be most highly expressed. 1-5 mg of total RNA was used in the mRNA isolation and the procedure carried out according to the manufacturer's instructions using Streptavidin MagneSphere® Paramagnetic Particles. The eluted mRNA was concentrated following the manufacturer's instructions and used for cDNA synthesis. A ThermoScript cDNA synthesis kit (Invitrogen) was used for making the cDNA with Oligo dT as the primer and the reaction carried out according to the manufacturer's instructions. The cDNA was used in PCRs with *DtFT* and *DtTSF* specific forward primers (Appendices A.1.5.3 and A.1.6.3) and Oligo dT as the reverse primer following the method set out in section 2.2.4. Product purification and sequencing was then undertaken as described in sections 2.2.5 and 2.2.6.

### **3.2.2. Functional complementation of Arabidopsis mutants**

Functional complementation of Arabidopsis mutants was undertaken for each gene of interest in this study and was carried out according to the method set out in section 2.2.8.3 and 2.2.9. The final step of the GATEWAY reaction ligated the isolated CDS for each gene into the pB2GW7 vector (Figure 3.1) which was then transformed into *E. Coli* for verification before transforming into *Agrobacterium* for floral dipping. For the analysis of transgenic plants in T<sub>1</sub>, T<sub>2</sub> and T<sub>3</sub> generations, leaf material was collected and DNA extracted using the CTAB buffer method (Stewart and Via, 1993) adapted for use with low amounts of starting material. The DNA was diluted to 100 ng/μl and used in PCR with primers designed to amplify the transgene. The products were visualised on a 1% (w/v) agarose (AGTC BioProducts)/1xTAE gel stained with GelRed (Biotium) at 100V for 20 minutes and placing on a UV lightbox (Syngene G-box with GeneSnap software). For T<sub>2</sub> generation of Arabidopsis complementation with *DtFLC* and T<sub>3</sub> generation of Arabidopsis complementation with *DtFT*, RNA was extracted from leaf material using a half volume Z6 buffer method (Logemann *et al.*, 1987). Real time PCR was then performed on these samples. *AtActin*, *AtTIP41* and *Atβ-tubulin* were used as housekeeping genes for normalisation and the standard was made from PCR products amplified from cDNA using real time primer pairs for the housekeeping genes, and

*DtFT* and *DtFLC* (Appendices A.2.1.1 and A.2.1.2). These were combined and purified using a PCR purification kit (Qiagen) before diluting from  $10^0$  to  $10^{-10}$ . The real time PCR was performed using  $S10^{-6}$  –  $S10^{-9}$  as laid out in section 2.2.10 with  $T_2$  FLC complementation samples run with iTaq™ Universal SYBR® Green Supermix (BioRad) and  $T_3$  FT complementation samples run with Go-Taq® qPCR master mix (Promega) with the addition of 10 nm (final concentration) Fluorescein calibration dye (Bio-Rad).



**Figure 3.1 – pB2GW7 plasmid used as binary vector in complementation experiments.**

Using the GATEWAY reaction, the isolated CDS for each gene was ligated into the region between attR1 and attR2, replacing the *ccdB* gene. Sm/SpR – Spectinomycin resistance gene, LB – Left border, Bar – BASTA resistance gene, p35S – 35S promoter, T35S – 35S terminator, RB – Right border (Karimi *et al.*, 2002).

### 3.2.3. Diurnal time course experiment

*Dt* SSD6 seed was sown onto F<sub>2</sub>S soil in p24s and topped with vermiculite. These were placed into a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C with 16 hours light, 8 hours dark. After 2.5 weeks, seedlings were transplanted into 5 inch pots of M2 soil and placed into two controlled environment cabinets (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C with 16 hours light, 8 hours dark. The cabinets were off-set from each other by 12 hours to enable harvesting during a normal working day. At four weeks

from sowing, the diurnal time course began. Over 52 hours, samples were collected from two sets of eight plants at 3 hour intervals beginning 30 minutes after the lights come on (ZT0). Material was collected from the newest fully expanded leaf on each plant.

#### **3.2.4. Developmental time course**

*Dt* SSD6 seed was sown into F<sub>2</sub>S soil and placed in a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C with a 16 hour photoperiod until initiation and bolting. Leaf material was collected at time point ZT15.5 once a week from sowing until two week post-initiation of bolting. Collecting material at the end of the long day enabled the levels of *CO* and *FT* to be measured.

#### **3.2.5. Gene expression analysis using real time PCR**

RNA was extracted from leaf material collected from both diurnal and developmental experiments using the Z6 buffer method (Logemann *et al.*, 1987) as laid out in section 2.2.2. RNA from each sample was DNase treated using an Ambion® Turbo DNA-free™ DNase kit and finally resuspended in 12 µl DEPC-treated H<sub>2</sub>O. 2 µg RNA was synthesised into cDNA using a ThermoScript cDNA synthesis kit (Invitrogen). Standards for the real time PCR were made using PCR products amplified from *Dt* cDNA using gene-specific real time primers for all six key genes and three housekeeping genes (*DtTIP41*, *DtCACs*, *Dta-tubulin*) (Appendix A.2.1.3). These were PCR purified using a PCR purification kit (Qiagen) and eluted in 20 µl H<sub>2</sub>O. Each was quantified and an equal concentration of each product was mixed together. This was diluted from 10<sup>0</sup> to 10<sup>-10</sup> and the real time PCR performed using S10<sup>-6</sup> – S10<sup>-9</sup>. Analysis was done according to section 2.2.10.

### **3.3. Results**

#### **3.3.1. CONSTANS (CO)**

##### **3.3.1.1. Isolation of a rocket *CO* (*DtCO*)**

Using PCR and RACE-RLM (Invitrogen), the full length CDS and gDNA sequence of the rocket *CO* gene was successfully isolated (Appendix B.1.1.1 to B.1.1.3) (Figure 3.2). It was found to have two large exons and a small intron. The CDS is 1035 nt long compared to Arabidopsis which is 1122 nt. The isolated full length

CDS for *DtCO*, when BLASTed, revealed that the nucleotide sequence has 88% identity to *B. napus CO* homologue (*Bn9CON10*) gene and 77% identity to *A. thaliana CO*. When the nucleotide sequence was translated using ExPASy (<http://web.expasy.org/translate/>) and BLASTed, the protein sequence was shown to have a 79% identity to CO homologue from *B. napus* (Figure 3.3) and 65% identity to *A. thaliana CO*. The DtCO protein contains two B-box motifs at the N-terminus and a CCT domain at the C-terminus like Arabidopsis.



**Figure 3.2 – Structure of *DtCO* gene from *Diplotaxis tenuifolia*.**

Red bars represent exons, black line represents non-coding regions, intron and UTR.

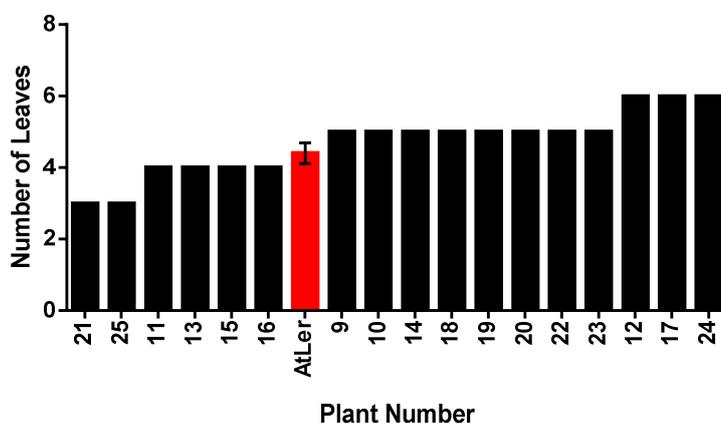
CONSTANS homolog [Brassica napus]						
Sequence ID: <a href="#">gb AAC27695.1 </a> Length: 366 Number of Matches: 1						
Range 1: 1 to 366 <a href="#">GenPept</a> <a href="#">Graphics</a> <span style="float: right;">▼ Next Match ▲ Previous Match</span>						
Score	Expect	Method	Identities	Positives	Gaps	
523 bits(1348)	0.0	Compositional matrix adjust.	290/367(79%)	309/367(84%)	24/367(6%)	
Query 1		MLKQETNNMSNRE---ARACDTCRSTICTVYCHADSAYLCNSCDAQVHSANRVASRHKR				56
Sbjct 1		M KQE+NN+ NRE ARACDTC STICTVYCHADSAYLCNSCDAQVHSANRVASRHKR				60
Query 57		VPVCECERAPAAAFMCEADDVSLCTACDEDVHSANSLAKRHRVPVVPVTGNSYTPLAST				116
Sbjct 61		V VCESCERAPAAAFMCEADDVSLCTACD +VHSAN LA+RH RVPVVP+ GNS + LA+ VRVCESCERAPAAAFMCEADDVSLCTACDLEVHSANPLARRHQVPPVPIIGNSCSSLATA				120
Query 117		QHTSVTEPEKRVVL--EDDKETASWLFPKNSDDHNNNNQNNELLFNDDYLDLADYNSSM				174
Sbjct 121		HT+VTEPEKRVVL ED KETASWLFPKNSD HNNNNQNNELLF+DDYLDLADYNSSM NHTTVTEPEKRVVLVQEDAKETASWLFPKNSDYHNNNNQNNELLFSDDYLDLADYNSSM				180
Query 175		DYKFPGQYNQHQHKQDCTVPQNTYGGDRVVPLKIEETRGNVHHKQHNITYGSSECHYNYN				234
Sbjct 181		DYKF QYNQ +HKQDC VP+ NY GDRVVPL++EETRGN+ +KQ NITYGSS YN N DYKFTSQYNQPRHKQDCIVPEKNYSGDRVVPLQLEETRGNLRNKQONITYGSSGQYNNN				240
Query 235		GSINHNAYNPSVETDFVPEPTR-DTTVSHQKTPKG-----NREARVRLRY				277
Sbjct 241		GSINHNAYNPS+ETDFVPE T DTTVSH KT KG +REARVRLRY GSINHNAYNPSMETDFVPEQTAPDPTTVSHPKTHKGKTAQLPEPLIQILSPMDREARVRLRY				300
Query 278		REKKRRRFEKTIRYASRKAYAERRPRINGRFKMSETEAEQEAFNRLMYDPTGYGIV				337
Sbjct 301		REKKRRRFEKTIRYASRKAYAERRPRINGRFKMSETE E+QE+N MLMY D TGYGIV REKKRRRFEKTIRYASRKAYAERRPRINGRFKMSETEVEDQEYNTMLMYCD-TGYGIV				359
Query 338		PSSYGQK 344				
Sbjct 360		PS YGQK 366				

**Figure 3.3 – *DtCO* protein BLASTP result. *DtCO* protein (Query) has been aligned to *B. napus CO* homologue (sbjct).**

Red line shows B-box domain, blue line shows CCT domain.

### 3.3.1.2. Complementation results

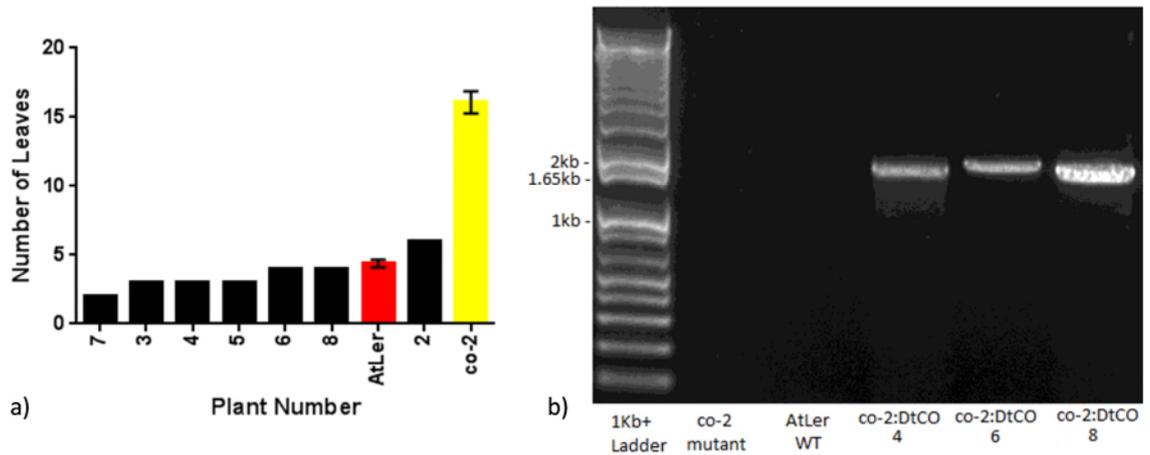
The isolated full length *DtCO* CDS under the control of the 35S promoter was transformed into *At Landsberg erecta (Ler)* to observe the effect of overexpression of *DtCO* in the wild type background, and transformed into *AtLer* background *co-2* mutant to see if *DtCO* was able to restore an early flowering phenotype. T<sub>1</sub> seed for both lines were sown and subjected to BASTA selection. Surviving plants were transplanted and bolting data collected. Figure 3.4 shows six plants flowered earlier than the wild type *AtLer* in LD, demonstrating the overexpression of *DtCO* can confer early flowering in *AtLer*. Transgene PCR was not carried out for *AtLer:DtCO* plants. *AtLer:DtCO11* was taken through to the T<sub>2</sub> generation as it flowered with fewer leaves than *AtLer* WT and produced a large amount of seed.



**Figure 3.4 – Flowering time for T<sub>1</sub> generation of *AtLer:DtCO* overexpression lines.**

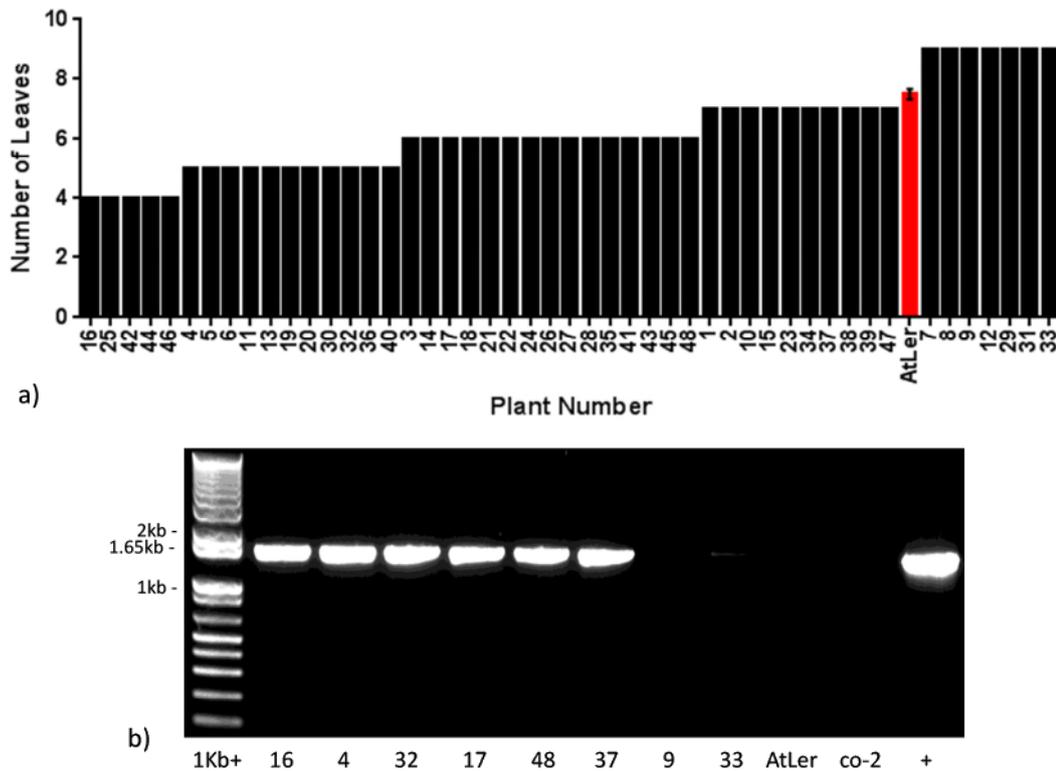
Histogram shows the number of leaves at a 1 cm bolt for *AtLer:DtCO*. *AtLer* WT bolting data is from a later date as these seeds did not germinate when sown in this experiment. Red bar shows mean number of leaves to bolt for *AtLer* wild type (mean  $4.4 \pm$  S.D 0.99 n=12). Error bars denote the standard error of the mean.

Figure 3.5a shows that complementation of the *co-2* mutant with *DtCO* restored the bolting phenotype to that of the wild type *AtLer* (red), or earlier. All plants that came through the BASTA selection were earlier flowering than the *co-2* mutant (yellow) by at least 30 days. The presence of the transgene in early flowering *co-2:DtCO* plants 4, 6 and 8, as well as *AtLer* and *co-2* was tested by PCR (Figure 3.5b) and showed that only the complementation lines contained the transgene. These three positive lines were used in the T<sub>2</sub> generation.



**Figure 3.5 – Flowering time and transgene presence for T<sub>1</sub> generation of *co-2:DtCO* complementation lines.**  
a) Histogram showing the number of leaves at a 1 cm bolt for *co-2:DtCO* transgenic lines. *AtLer* WT bolting data is from a later date as these seeds did not germinate when sown in this experiment. Red bar shows mean number of leaves to bolt for *AtLer* wild type (mean  $4.4 \pm S.D 0.99$  n=12). Yellow bar shows mean number of leaves to bolt for *co-2* mutant (mean  $16.1 \pm 2.3$  n=8). Error bars denote the standard error of the mean. b) Gel picture shows PCR detection of *DtCO* transgene presence in *co-2* mutant, *AtLer* wild type and *co-2:DtCO* plants 4, 6 and 8. These were chosen to continue onto T<sub>2</sub> generation.

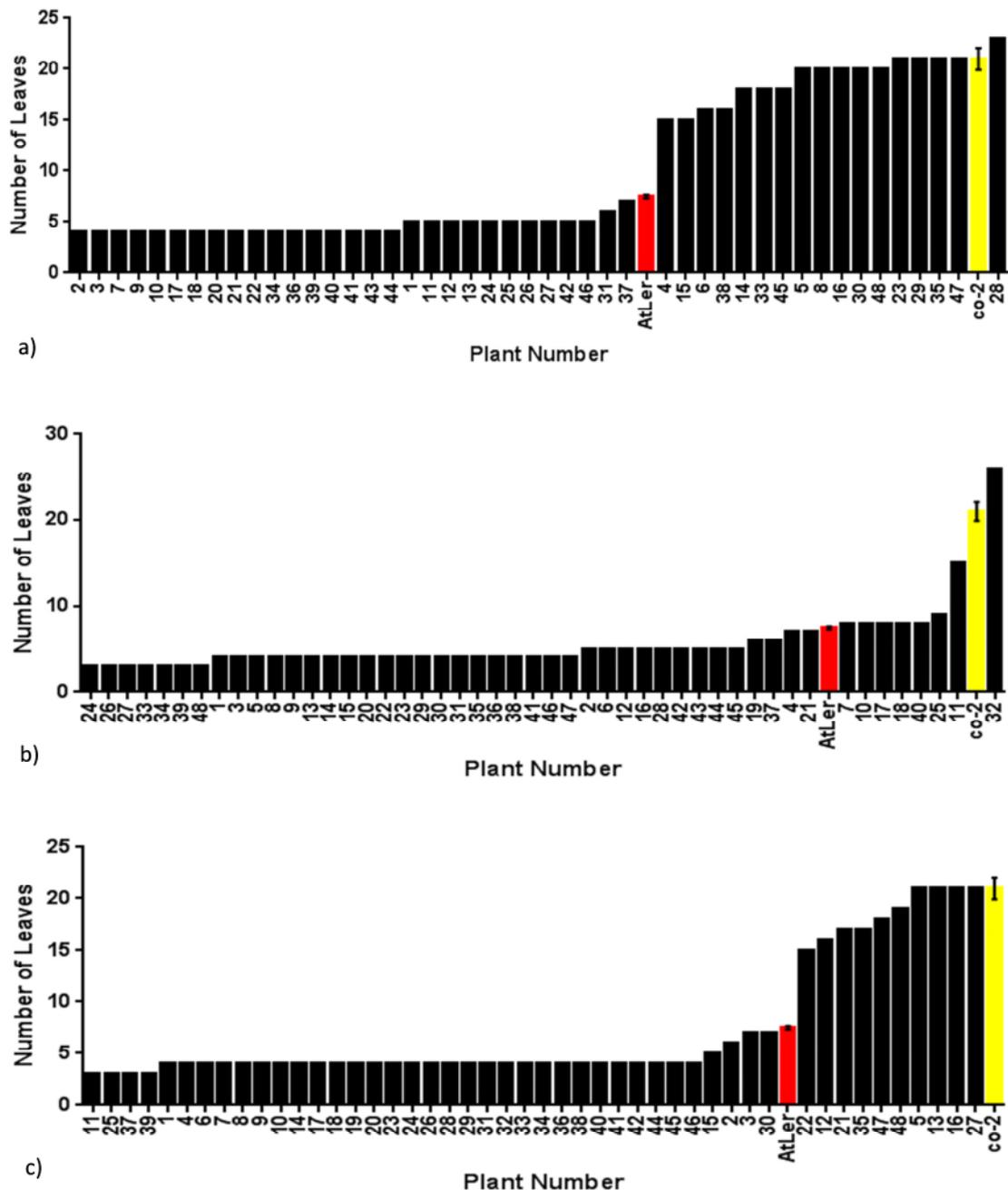
In the T<sub>2</sub> generation of *AtLer:DtCO* and *co-2:DtCO*, 48 seeds of each line were sown into Arabidopsis mix soil in p24s along with 24 seeds of each *AtLer* wild type and *co-2* mutant. No selection method was used, but flowering time by rosette leaf number at 1 cm bolt was recorded. The T<sub>2</sub> generation is expected to be segregating in a 3:1 ratio of early bolting plants:wild type bolting plants if a single copy of the transgene is present. In line *AtLer:DtCO 11* a 3:1 ratio was observed (Figure 3.6a) where three out of four plants were earlier bolting by at least three days compared to the *AtLer* wild type (red bar). DNA was extracted and the transgene presence tested by PCR for several plants in the line. *AtLer:DtCO 11* plants 16, 4, 32, 17, 48 and 37 all bolted with fewer leaves than *AtLer* WT, and at least three days earlier and were expected to contain the transgene. *AtLer:DtCO 11* plants 9 and 33 bolted with more leaves than the *AtLer* WT and were not expected to contain the transgene. The transgene PCR shows these to be a correct assumption (Figure 3.6b).



**Figure 3.6 – Flowering time data and transgene amplification for *AtLer:DtCO 11* T<sub>2</sub> line.**

a) Flowering time data for line *AtLer:DtCO 11*. Rosette leaf number was recorded when bolt reached 1 cm. Red bar shows mean number of leaves at flowering for *AtLer* wild type (mean  $7.5 \pm 0.8$  n=24). Error bars denote the standard error of the mean. Chi-squared tests assessing goodness of fit for data to an expected 3:1 early:late bolting phenotype show that *AtLer:DtCO 11* was not significantly different to the 3:1 ratio ( $p < 0.05$ ). b) Gel pictures showing the amplification of the transgene by PCR for three T<sub>2</sub> generation lines. Bands present at 1.65 kb show where transgene is present, absence of the band at 1.65 kb shows no transgene is present. 1 kb+ ladder used. Positive control (+) was generated using DNA extracted from a plasmid known to contain the *CO* transgene.

The *co-2:DtCO* T<sub>2</sub> generation was sown in the same way as the *AtLer:DtCO* T<sub>2</sub>. Flowering time data were recorded for all plants (Figure 3.7). The *co-2* mutant plants flowered late with an average of 21 leaves (yellow bar) and 14 days later than the *AtLer* WT, so where segregation is occurring, the ratio should be 3:1 early:late flowering. A Chi-squared goodness of fit test was applied to the flowering data to assess how well each fitted the expected 3:1 ratio. All lines *co-2:DtCO 4, 6* and *8* showed a ratio of 3:1. This suggests segregation of a single transgene is occurring in these populations of a single transformation event and that *DtCO* is able to complement the function of *AtCO*.

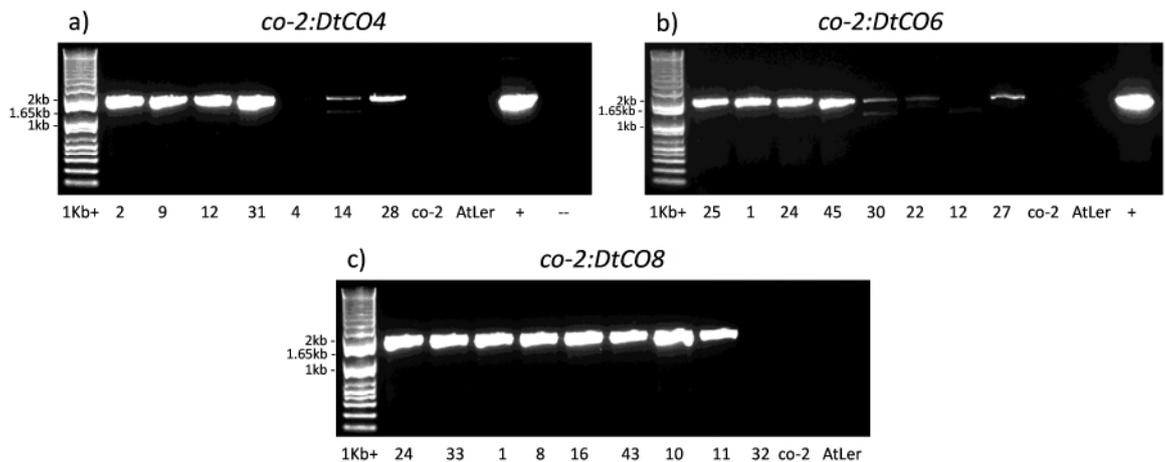


**Figure 3.7 – Flowering time data for *co-2:DtCO* T<sub>2</sub> lines.**

Rosette leaf number was recorded when bolt reached 1 cm. a) Flowering time data for line *co-2:DtCO 4*, b) Flowering time data for line *co-2:DtCO 6*, c) Flowering time data for line *co-2:DtCO 8*. Red bar shows mean number of leaves at flowering for *AtLer* wild type (mean  $7.5 \pm 0.8$  n=24). Yellow bar shows mean number of leaves at flowering for *co-2* mutant (mean  $21.1 \pm 5.2$  n=23). Error bars denote the standard error of the mean. A Chi-squared test assessing goodness of fit for data to an expected 3:1 early:late bolting phenotype shows that no lines were significantly different to the 3:1 ratio ( $p < 0.05$ ) where 7 leaves was the cut off for early flowering.

DNA was extracted from some lines and the transgene PCR carried out in the same way as the T<sub>1</sub> generation. The results reveal that the transgene is present in all early flowering plants in *co-2:DtCO 4*, *co-2:DtCO 6* and *co-2:DtCO 8* (Figure 3.8).

However, the results for the plants flowering with a similar number of leaves to *AtLer* wild type or more were mixed. Several plants did not contain the transgene (*co-2:DtCO 4\_4*, *co-2:DtCO 6\_12* and *co-2:DtCO 8\_32*) but in other plants, a faint band was observed where there was some amplification of the transgene (*co-2:DtCO 4\_14 and \_28* and *co-2:DtCO 6\_30, \_22 and \_27*) which might be due to contamination in the PCR or DNA samples, or where the transgene is present but poorly expressed, perhaps due to a positional effect.

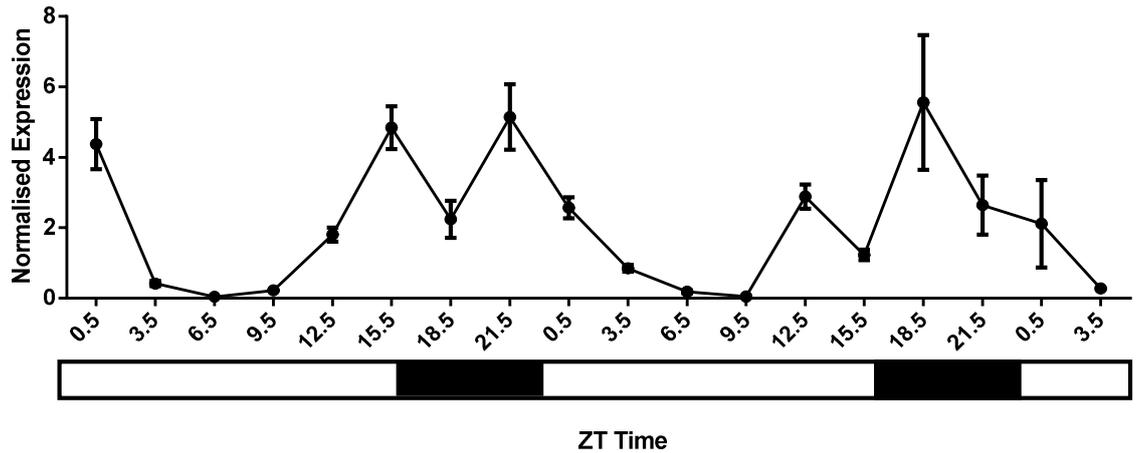


**Figure 3.8 – Transgene amplification in complementation line *co-2:DtCO T<sub>2</sub>*.**

Gel pictures showing the amplification of the transgene by PCR for three  $T_2$  generation lines. a) Line *co-2:DtCO 4* b) Line *co-2:DtCO 6* c) Line *co-2:DtCO 8*. Bands present at 1.65 kb show where transgene is present, absence of the band at 1.65 kb shows no transgene is present. 1 kb+ ladder used. Positive control (+) was generated using DNA extracted from a plasmid known to contain the *DtCO* transgene, the negative control (-) contained no DNA.

### 3.3.1.3. Diurnal time course experiment

RNA was extracted from leaf material collected at 3 hour intervals from plants grown in LDs over a 52 hour period. This was used to synthesise cDNA and real time PCR was used to analyse gene expression over the time course. Three housekeeping genes, *DtTIP41*, *DtCACS*, *Dta-tubulin* were used to normalise the expression of the gene of interest. The expression of *DtCO* fluctuates over 24 hours in a regular pattern. Expression is high at ZT0.5 and decreases during the light period until ZT9.5 where it begins to increase and remains at a high level until the end of the dark period (Figure 3.9). The data suggests that *DtCO* is diurnally regulated.

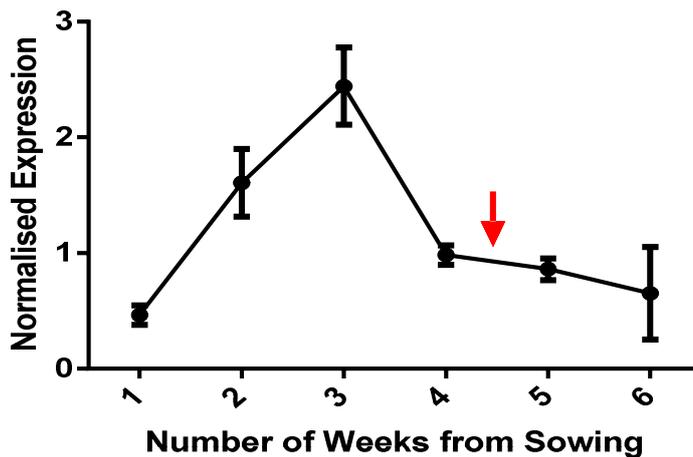


**Figure 3.9 – Diurnal Time Course of *DtCO*.**

Expression of *DtCO* showing the diurnal changes in expression levels. Lights providing a LD 16h photoperiod were switched on at ZT0 (white bar) and off at ZT16 (black bar). Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars show the standard error of three technical replicates.

### 3.3.1.4. Developmental time course

Real time PCR was performed using cDNA made from RNA extracted from leaf material collected at the end of the LD photoperiod (ZT15.5). Three housekeeping genes, *DtTIP41*, *DtCACS* and *Dta-tubulin* were used to normalise the data for each gene of interest. Figure 3.10 shows the expression of *DtCO* increasing during early development, before falling as the transition to the reproductive phase begins and staying low once bolting is initiated.



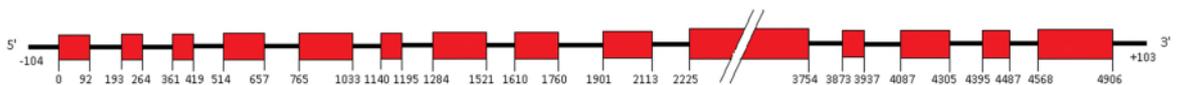
**Figure 3.10 – Expression of *DtCO* across development.**

Leaf material was collected at ZT15.5. The red arrow indicates where the plants initiated bolting. Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars denote the standard error of three technical replicates.

### 3.3.2. GIGANTEA (GI)

#### 3.3.2.1. Isolation of a rocket *GI* (*DtGI*)

The *DtGI* gene was isolated from both cDNA and gDNA using Kod Hot start and Kod Xtreme™ (Novagen®) PCR kit. The full length gene products were cloned using Zeroblunt® TOPO® PCR cloning kit (Invitrogen) due to the large size of the gene, and clones were sequenced. This showed that the CDS was 3540 nt and the full length genomic DNA sequence was 4906 nt excluding the 5' and 3' UTR (Figure 3.11, Appendices B.1.1.9 and B.1.1.10). The gene structure has 14 exons and 13 introns as observed in Arabidopsis but the CDS is 3522 nt. The full length CDS for *DtGI* has a 92% identity to a predicted *B. rapa GI* gene, and 89% identity to *A. thaliana GI*. The CDS was translated and the DtGI protein was found to be 1179 amino acids in length (Appendix B.1.1.11) compared to Arabidopsis which is 1173 amino acids long. When BlastP was used, the *DtGI* protein had a 94% identity to *B. rapa GI* (Figure 3.12) and 91% identity to *A. thaliana GI*.



**Figure 3.11 – Structure of *DtGI* gene CDS from *Diplotaxis tenuifolia*.**

Red bars represent exons, black lines represent introns and UTR.

protein GIGANTEA [Brassica rapa]  
Sequence ID: [ref|NP\\_001288824.1](#) Length: 1183 Number of Matches: 1  
[▶ See 1 more title\(s\)](#)

Range 1: 1 to 1183 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
2174 bits(5633)	0.0	Compositional matrix adjust.	1126/1196(94%)	1147/1196(95%)	30/1196(2%)
Query 1		MTSPTSSEKWTDLQFSSLLWSPPRDPQQHKDQVVAYVEYFGQFTSEQFPDDIAELVRNQ			60
Sbjct 1		MTSPTSSE+WTDLQFSSLLWSPPRDPQQHKDQVVAYVEYFGQFTSEQFPDDIAELVRNQ			60
Query 61		YPSTEKRLDDVLFAMFVLHHPHGHAVILPIISCLIDGSLVYSKEAHPFASFISLVSPNS			120
Sbjct 61		YPSTEKRLDDVLFAMFVLHHPHGHAVILPIISCLIDG+LVYSKEAHPFASFISLVSPNS			120
Query 121		ENDYSEQWALACGEILRILTHYNRPIYKREQQNGDTEKASTSGSGSPTLSEKVASPGQ			180
Sbjct 121		ENDYSEQWALACGEILRILTHYNRPIYK EQQNG+TESKASTSGS PTLSE K SPGQ			178
Query 181		NERKPLRPLSPWISDILLAAPLGRISDYFRWCSGVMGKYAAGELKPPTI-----			229
Sbjct 179		HERKPLRPLSPWISDILLAAPLGRISDYFRWCSGVMGKYAAGELKPPTI			238
Query 230		-ASRSGKHPQYMPSTPRWAVANGAGVILSVCDDEVARYETATLTAVAVPALLLPPTTS			288
Sbjct 239		++RSGKHPQYMPSTPRWAVANGAGVILSVCDDEVARYETATLTAVAVPALLLPPTTS			298
Query 289		LDEHLVAGLPALEPYARLFHRYAIAIATPSATQRLLLGLEAPPSWAPDALDAAVQLVELL			348
Sbjct 299		LDEHLVAGLPALEPYARLFHRYAIAIATPSATQRLLLGLEAPPSWAPDALDAAVQLVELL			358
Query 349		RATEDYASGVRLPRNWMHLHFLRAIGIAMSMRAGVAADAAAALLFRILSQPALLFPPLSQ			408
Sbjct 359		RA EDYASGVRLPRNWMHLHFLRAIGIAMSMRAGVAADAAAALLFRILSQPALLFPPLSQ			418
Query 409		AEGVEIQNSPLGGYSSNHRTOAEVSAEATIEATAQGIASMLCAHGPEVEWRICTIWEAA			468
Sbjct 419		AEGVEI+++P+GGY SN+R Q EV AAEATIEATAQGIASMLCAHGPEVEWRICTIWEAA			478
Query 469		YGLIPLNSSAVDLPEIIVATPLQPPILSWNLYIPLLKVLEYLPRGSPSEACLMKIFVATV			528
Sbjct 479		YGLIPLNSSAVDLPEIIVATPLQPPILSWNLYIPLLKVLEYLPRGSPSEACLMKIFVATV			538
Query 529		ETILSRTFPPETSIRKAR+SL TR SATKNLAMAELRAMVHALFLESCAGVEIASRL			587
Sbjct 539		ETILSRTFPPETS---IRKARASLATRSSATKNLAMAELRAMVHALFLESCAGVEIASRL			595
Query 588		LFVVLTVCVSHEAQSSGSKRRRSEDASSTAEENQHVSDNQTSNRKSRSVKGGPAAAFDS			647
Sbjct 596		LFVVLTVCVSHEAQSSGSKRRRSE+ +TAEENQ DNQTSNRKSR+VKGGP AAFDS			651
Query 648		YVLAAVCALACEVQLYPMISGGGNFNSAVAATITKSVKINGSSNEYGSGIDSIAIKHTRR			707
Sbjct 652		YVLAAVCALACEVQLYPMISGGGNFNSAVAATITKSVKINGSSNEYG+GIDSIAIKHTRR			711
Query 708		ILAILEALFSLKPSVGTWPWSYSSSEIVAAAMVAAHISELFRRSKALTHALSGLMRCKWD			767
Sbjct 712		ILAILEALFSLKPSVGTWPWSYSSSEIVAAAMVAAHISELFRRSKALTHALSGLMRCKWD			771
Query 768		KEIHKRASSLYNLIDVHSKVVASIVDKAEPLAAYLKNAPLQKDALTCNLNWKQNTS----			823
Sbjct 772		KEIHKRASSLYNLIDVHSKVVASIVDKAEPLAAYLKNAP+QKD+L C+NWKQN +			831
Query 824		GFGTAAVMSTSRNEMTPRGGNNHKCARHSDGSGSRSSSSSEKGIKDLLLDASDLANFLTA			883
Sbjct 832		GFGTAAV TSTRNEM PRGG NHK ARHSDGSGSR SSS+KGIKDLLLDASDLANFLTA			889
Query 884		DRLAGFYRGTQVLLRSILAEPKPELSFVSVLLWHKLIASPEIQPTVESTSAQQGWRQVVD			943
Sbjct 890		DRLAGFYRGTQVLLRSILAEPKPELSFVSVLLWHKLIASPEIQPT ESTSAQQGWRQVVD			949
Query 944		ALCNVVSATPAKAAAAVVLAERELQPWIAKDDEEGQKMWKINQRIVKVMVELMRNHRP			1003
Sbjct 950		ALCNVVSATPAKAAAAVVLAERELQPWIAKDDEEGQKMWKINQRIVKVMVELMRNHRP			1009
Query 1004		ESLVIXASADLLLRATDGMVDGEACTLPQLELLEATARAIPVLAWGPSSGLAVVDGLS			1063
Sbjct 1010		ESLVI ASADLLLRATDGMVDGEACTLPQLELLEATARAIPVLAWGPSSGLAVVDGLS			1069
Query 1064		NLLKCRLPATIRCLSHPSAHVRLSTSVLRDIMNQSGITATTKATPKPPPTITTEKNGT			1123
Sbjct 1070		NLLKCRLPATIRCLSHPSAHVRLSTSVLRDIMNQSGIT TTKATPKPPPTITTEKNGT			1127
Query 1124		DSPSYRFFNAAAIIDWKADIQKCLNWEAHSLLSTTMPTQFLDTAARELGCTISMSSQ			1179
Sbjct 1128		DSPSYRFFNAAAIIDWKADIQKCLNWEAHSLLSTTMPTQFLDTAARELGCTISMSSQ			1183

**Figure 3.12 – DtGI protein BLASTP result.**

DtGI protein (Query) has been aligned to *Brassica rapa* GI homologue (sbjct).

### 3.3.2.2. Complementation results

The full length *DtGI* CDS was fully sequenced and checked before it was cloned into the transformation vector pB2GW7 using GATEWAY. This vector containing the *DtGI* sequence was transformed into *A. tumefaciens* and floral dipping was done on both *AtCol0* and *gi-3* mutant plants. Seed was recovered from the dipped plants and sown out for selection using BASTA (section 3.2.2). Two attempts were made with both *AtCol0:DtGI* and *gi-3:DtGI* seed but unfortunately no plants came through the BASTA treatment, so no *DtGI* transformants were obtained.

### 3.3.2.3. Diurnal time course experiment results

The change in gene expression of *DtGI* was investigated in LD over a 52 hour time course. Expression is low at dawn and rises steadily until it peaks at ZT9-10. Expression then falls to very low levels by the beginning of the night where it remains until dawn (Figure 3.13). This expression pattern suggests that *DtGI* is diurnally regulated.

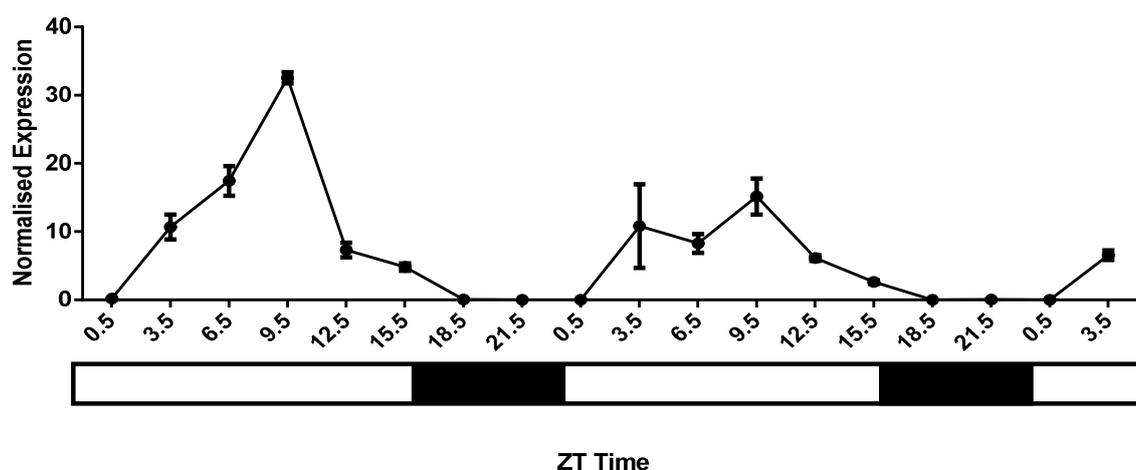
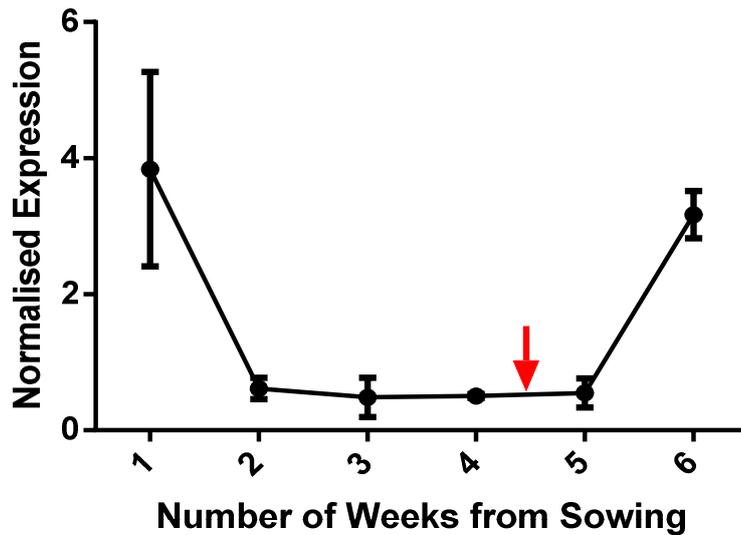


Figure 3.13 – Diurnal Time Course of *DtGI*.

Expression of *DtGI* showing the diurnal changes in expression levels. Lights were switched on at ZT0 (white bar) and off at ZT16 (black bar). Expression was normalised to the average expression of *Dta-tubulin*, *DtCAC5* and *DtTIP41*. Error bars show the standard error of three technical replicates.

### 3.3.2.4. Developmental time course

Leaf material was collected at weekly intervals at ZT15.5 over the experimental period. RNA was extracted and cDNA made from each sample. Real time PCR was carried out on each sample to examine the change in gene expression over time. Figure 3.14 shows that the levels of *DtGI* expression start and finish high, but that during most of development the levels are low.



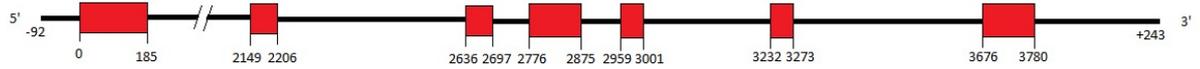
**Figure 3.14 – Expression of *DtGI* across development.**

Leaf material was collected at ZT15.5. The red arrow indicates where the plants initiated bolting. Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars denote the standard error of three technical replicates.

### 3.3.3. FLOWERING LOCUS C (FLC)

#### 3.3.3.1. Isolation of a rocket *FLC* (*DtFLC*)

The *DtFLC* gene CDS was easily isolated using PCR with degenerate primers followed by GeneRacer™ Kit (RLM-RACE) (Invitrogen). The full length gDNA product was large and so isolated using Kod Xtreme™ (Novagen®) PCR kit and sequenced. The isolation of the *DtFLC* gene revealed that it contained 7 exons and 6 introns (Figure 3.15). The CDS is 594 nt long and the gDNA (excluding 5' and 3' UTR) is 3780 nt long (Appendices B.1.1.4 and B.1.1.5). The full length CDS for *DtFLC* was BLASTed and revealed that the nucleotide sequence has 94% identity to *Sinapis alba FLC*, *B. nigra FLC* and *B. napus FLC.A10*. It was also found to have 78% identity to *A. thaliana FLC1-like* and *MAF1*. When translated, the protein sequence has a 95% identity to *S. alba FLC*, 94% identity to *B. nigra FLC* and 84% identity to *A. thaliana FLC*. The *DtFLC* protein contains the MADS box at the N-terminus and a K-box domain in the middle of the sequence (Figure 3.16, Appendix B.1.1.6).



**Figure 3.15 – Structure of *DtFLC* gene CDS from *Diplotaxis tenuifolia*.**  
Red bars represent exons. Black lines represent introns and UTR.

flowering locus C protein [Sinapis alba]  
Sequence ID: [gb|ABP96967.1](#) Length: 197 Number of Matches: 1

Range 1: 1 to 197 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

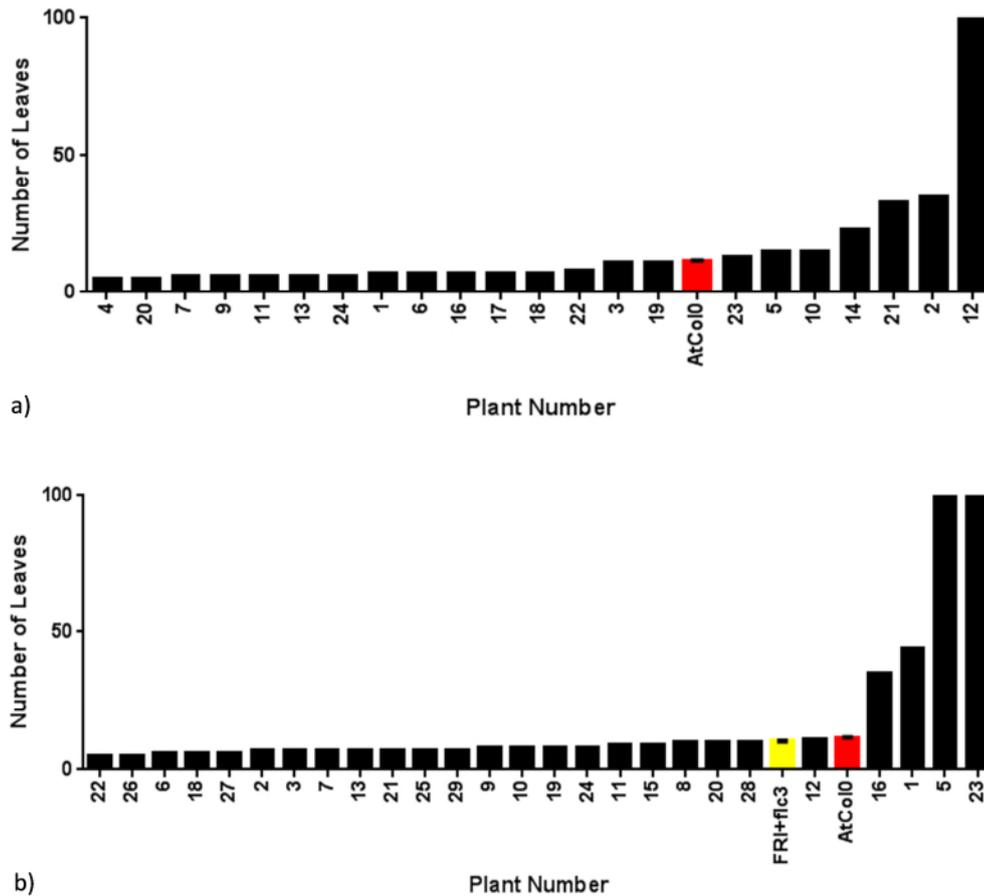
Score	Expect	Method	Identities	Positives	Gaps
333 bits(854)	3e-113	Compositional matrix adjust.	187/197(95%)	191/197(96%)	0/197(0%)
Query 1	MGRKKLEIKLIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSSSSG				60
Sbjct 1	MGRKKLEIK IENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYS SSG				60
Query 61	DNLVKICDRYGKQHDDDLKALDRQSKPLNCGSHHELLELVESKLEESNVDNVS VGSILVQL				120
Sbjct 61	DNLVKI DRYGKQHDDDLKALDRQSKPLNCGSHHELLELVESKLEESN +NVSVGSILVQL				120
Query 121	EEHLENALS VTRARKTEMLKLVENLKEKEKLEENHVLASQMEKNNLRAEADDM EVS				180
Sbjct 121	EENLESALS VTRARKTEMLKLVENLKEKEKLEENHVLASQMEKNNL RAEAD+MEVS				180
Query 181	PGQISDINLRVTLPLLN 197				
Sbjct 181	PGQISDINL VTLPLLN 197				

**Figure 3.16 – *DtFLC* protein BLASTP result. *DtFLC* protein (Query) has been aligned to *Sinapis alba* FLC homologue (sbjct).**

Red line represents the MADS box domain and the blue line represents the K-box motif.

### 3.3.3.2. Complementation results

The isolated *DtFLC* CDS was transformed into *A.thaliana* wild type Col0 and *A.thaliana* mutant line *FRI<sup>+</sup>flc3 null*. T<sub>1</sub> seed was selected using BASTA and transformants transplanted and the number of rosette leaves counted at flowering. The T<sub>1</sub> generation for the *AtCol0:DtFLC* shows that there is a range of flowering times across the population, with some plants which flower with fewer leaves, but in a similar number of days than the *AtCol0* wild type (31 days to flower) (*AtCol0:DtFLC 4, 20*) and some which are very late (93 days to flower) with a high number of leaves (*AtCol0:DtFLC 12*) (Figure 3.17a). Figure 3.17b shows the flowering time data of the T<sub>1</sub> generation of *FRI<sup>+</sup>flc3:DtFLC* expression lines. Here, all but four plants flower at a very similar number of days and leaves to the *FRI<sup>+</sup>flc3* mutant or *AtCol0* wild type plants. Unfortunately, when transgene PCR was carried out on DNA from the T<sub>1</sub> generation plants, the results showed that the samples were cross-contaminated, as there was amplification of the transgene in the *FRI<sup>+</sup>flc3* mutant and *AtCol0* wild type DNA.

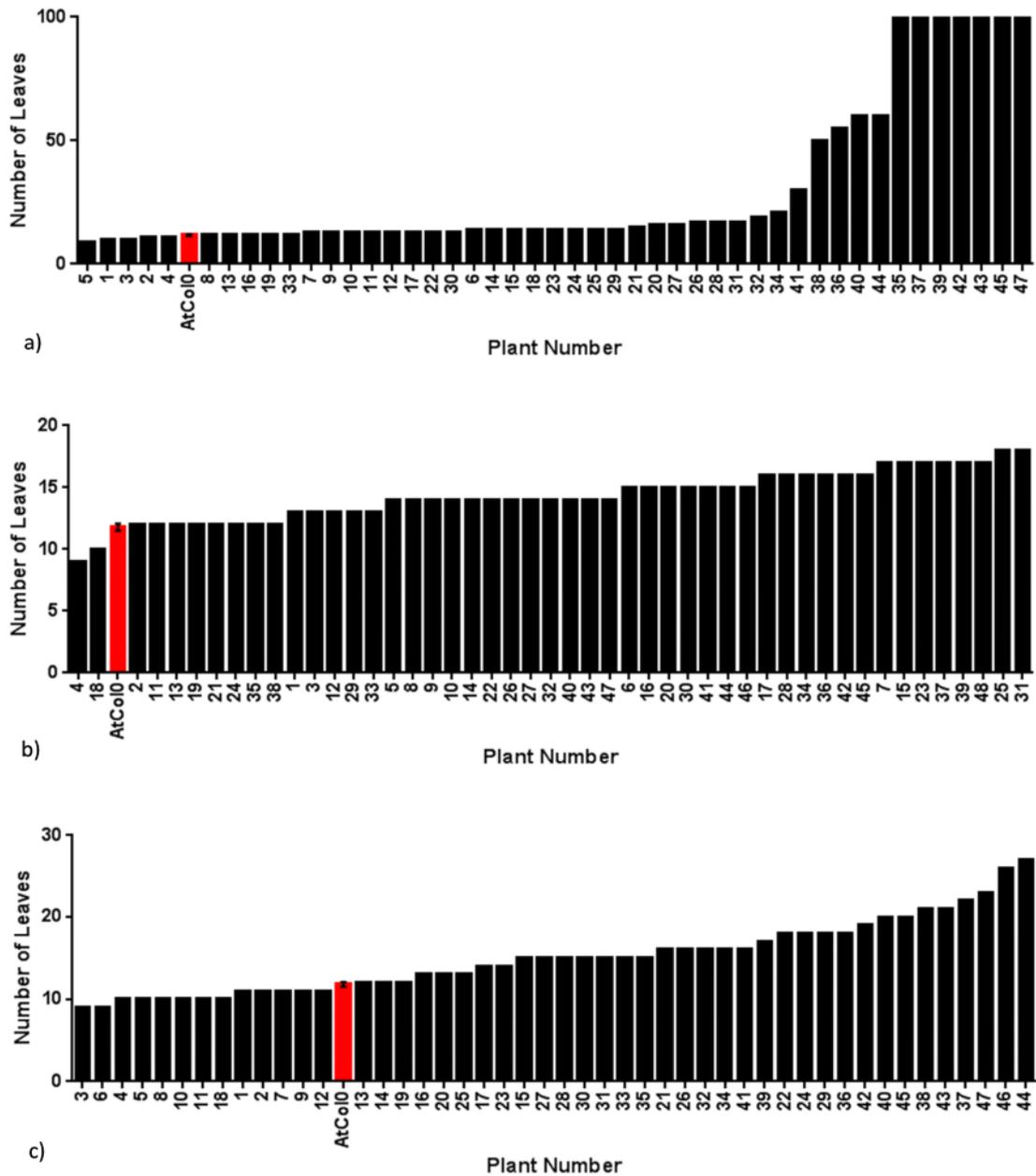


**Figure 3.17 – Flowering time of T<sub>1</sub> generation of *AtCol0:DtFLC* and *FRI<sup>+</sup>flc3:DtFLC* transformed plants.**

a) Flowering time recorded by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *AtCol0:DtFLC* overexpression lines. b) Flowering time recorded by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *FRI<sup>+</sup>flc3:DtFLC* expression lines. Red bar shows the *AtCol0* wild type flowering time data (mean  $11.7 \pm 1.5$  n=20) and yellow bar shows the *FRI<sup>+</sup>flc3* mutant flowering time data (mean  $10.4 \pm 1.3$  n=10).

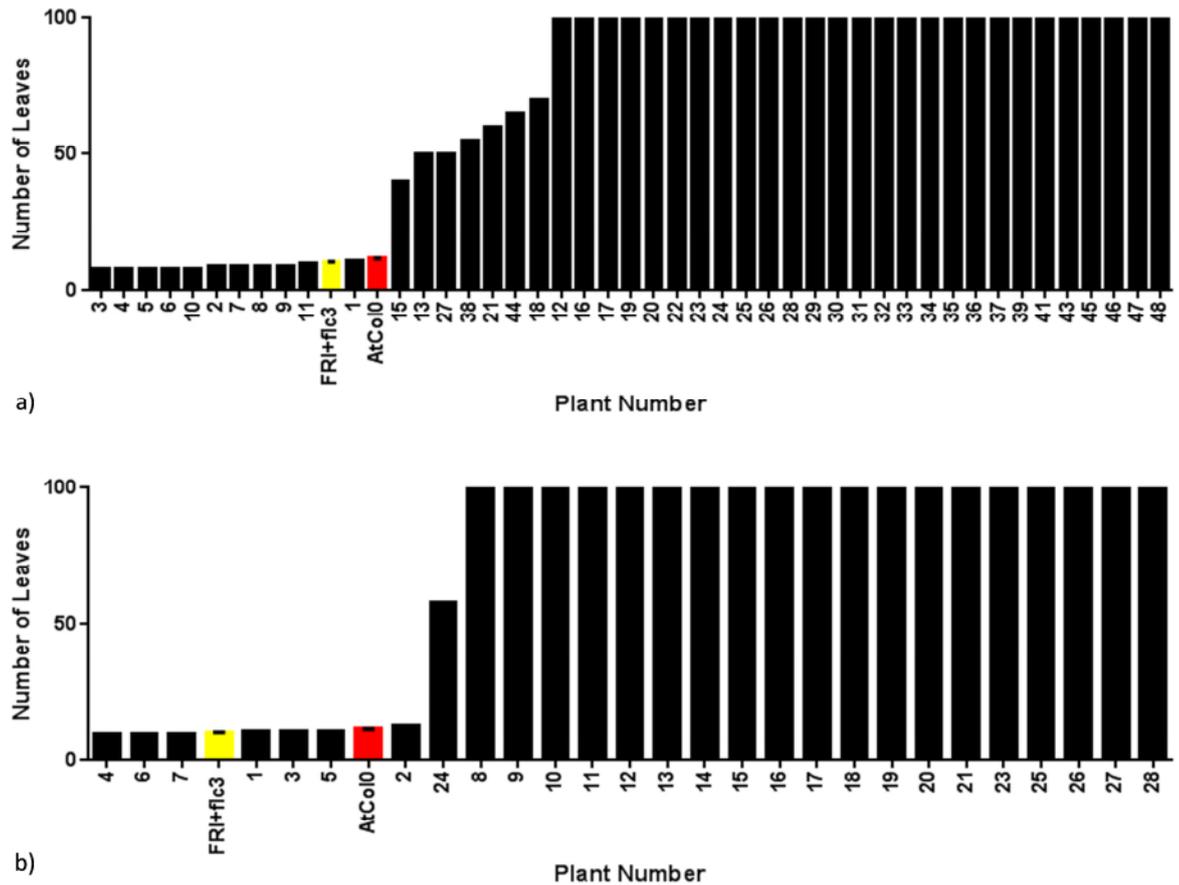
Seed was collected and sown for *AtCol0:DtFLC* 12, *AtCol0:DtFLC* 14, *AtCol0:DtFLC* 21, *FRI<sup>+</sup>flc3:DtFLC* 5 and *FRI<sup>+</sup>flc3:DtFLC* 23 as these displayed a late flowering phenotype in the T<sub>1</sub>. This gave rise to the T<sub>2</sub> generation and 48 plants were screened for their flowering time phenotype. In the *AtCol0* transformed lines, one of the three T<sub>2</sub> lines (*AtCol0:DtFLC* 21) displayed the expected 3:1 late:early flowering phenotype ratio according to the Chi-squared goodness of fit test (Figure 3.18c). Line *AtCol0:DtFLC* 12 had a larger number of wild type flowering plants than expected which was surprising due to the late flowering phenotype of the T<sub>1</sub> plant. Lines *AtCol0:DtFLC* 12 and *AtCol0:DtFLC* 14 were found to be significantly different to the 3:1 late:early flowering ratio (Figure 3.18a and Figure 3.18b). This suggests there may be multiple insertions of the transgene. In the *FRI<sup>+</sup>flc3:DtFLC*

complementation lines, the T<sub>2</sub> generation shows no significant difference to the 3:1 late:early flowering phenotype for lines FRI<sup>+</sup>*flc3:DtFLC* 5 (Figure 3.19a) and FRI<sup>+</sup>*flc3:DtFLC* 23 (Figure 3.19b). Overall the data suggest that *DtFLC* is able to complement the function of *AtFLC*. As for the T<sub>1</sub> generation, the transgene PCR for T<sub>2</sub> plants was found to be contaminated.



**Figure 3.18 – Flowering time data for *AtCol0:DtFLC* T<sub>2</sub> lines.**

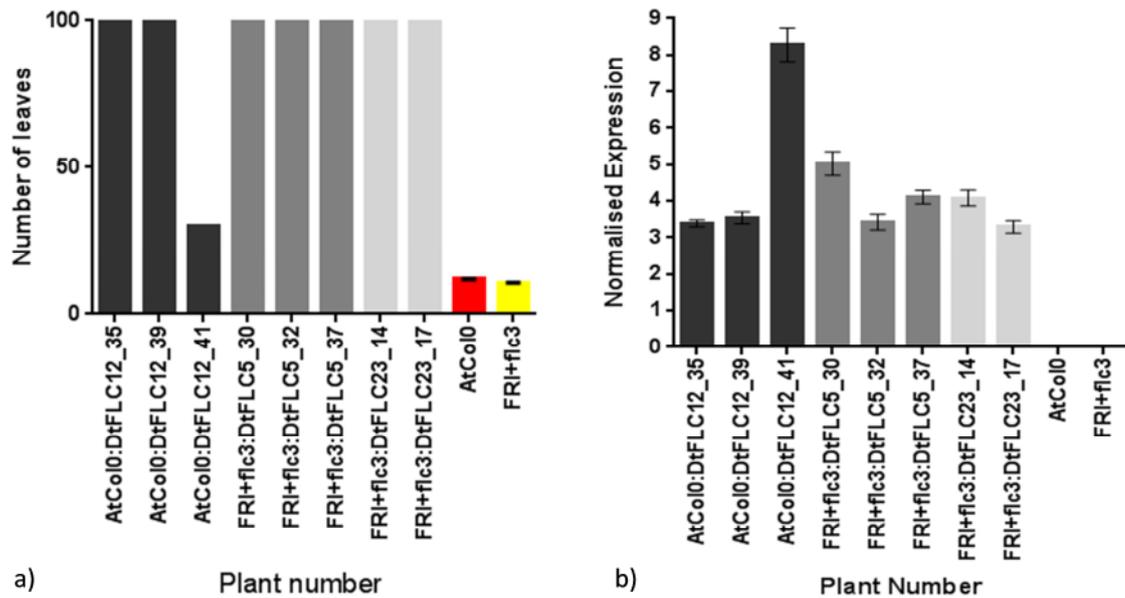
Rosette leaf number was recorded when bolt reached 1 cm. a) Flowering time data for line *AtCol0:DtFLC 12*, b) Flowering time data for line *AtCol0:DtFLC 14*, c) Flowering time data for line *AtCol0:DtFLC 21*. Red bar shows mean number of leaves at flowering for *AtCol0* wild type (mean  $11.8 \pm 1.5$  n=24). Error bars denote the standard error of the mean. Chi-squared tests assessing goodness of fit for data to an expected 3:1 late:early bolting phenotype show that line *AtCol0:DtFLC 21* was not significantly different to the 3:1 ratio ( $p < 0.05$ ). Lines *AtCol0:DtFLC 12* and *AtCol0:DtFLC 14* were found to be significantly different to the 3:1 ratio at  $p < 0.05$  if the cut off for early bolting is 11 leaves.



**Figure 3.19 – Flowering time data for FRI<sup>+</sup>flc3:DtFLC T<sub>2</sub> lines.**

Rosette leaf number was recorded when bolt reached 1 cm. a) Flowering time data for line FRI<sup>+</sup>flc3:DtFLC 5, b) Flowering time data for line FRI<sup>+</sup>flc3:DtFLC 23. Red bar shows mean number of leaves at flowering for *AtCol0* wild type (mean  $11.8 \pm 1.5$  n=24). Yellow bar shows mean number of leaves at flowering for FRI<sup>+</sup>flc3 mutant (mean  $10.7 \pm 1.2$  n=21). Black bars at 100 leaves are where plants did not bolt before the end of the experiment. Error bars denote the standard error of the mean. Chi-squared tests assessing goodness of fit for data to an expected 3:1 late:early bolting phenotype show that lines FRI<sup>+</sup>flc3:DtFLC 5 and FRI<sup>+</sup>flc3:DtFLC 23 were not significantly different to the 3:1 ratio ( $p < 0.05$ ).

RNA was extracted from a few plants in the T<sub>2</sub> generation lines *AtCol0:DtFLC 12*, FRI<sup>+</sup>flc3:DtFLC 5 and FRI<sup>+</sup>flc3:DtFLC 23. cDNA was synthesised and real time PCR used to investigate the expression of the transgene present in the transformed lines. Figure 3.20a shows the number of leaves at flowering for the lines of interest. Figure 3.20b shows the transgene expression in each of the plants. Interestingly, the transgene expression is highest in the plant that flowers with the fewest leaves. There is no transgene expression in the *AtCol0* wild type and FRI<sup>+</sup>flc3 mutant plants so there is no contamination in this experiment.

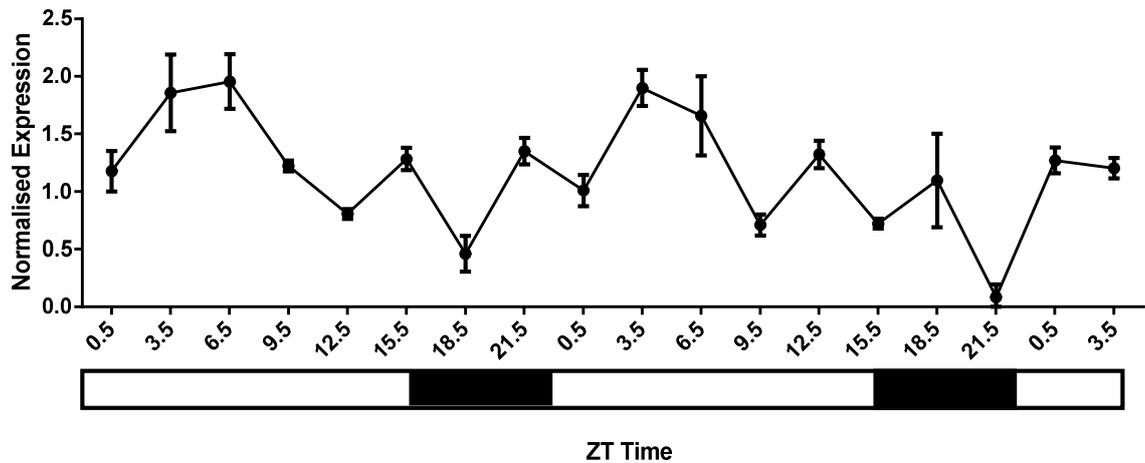


**Figure 3.20 – Flowering time and transgene expression of T<sub>2</sub> generation plants from lines *AtCol0:DtFLC 12*, *FRI+flc3:DtFLC 5* and *FRI+flc3:DtFLC 23*.**

a) Number of leaves at 1 cm bolt of T<sub>2</sub> generation plants from lines *AtCol0:DtFLC 12* (black bars), *FRI+flc3:DtFLC 5* (dark grey bars) and *FRI+flc3:DtFLC 23* (light grey bars). Bars with 100 leaves did not bolt before the end of the experiment. Error bars denote standard error of the mean. b) Transgene expression analysis using real time PCR for T<sub>2</sub> generation plants of lines *AtCol0:DtFLC 12* (black bars), *FRI+flc3:DtFLC 5* (dark grey bars) and *FRI+flc3:DtFLC 23* (light grey bars). Expression was normalised to the average expression of *Atβ-tubulin*, *AtActin2* and *AtTIP41*. Error bars denote standard error of three technical replicates.

### 3.3.3.3. Diurnal time course experiment results

The diurnal change in expression of *DtFLC* was observed using a diurnal LD time course where samples were taken every three hours over 52 hours. Figure 3.21 shows that there is little change in *DtFLC* expression over each 24 hour time period as the expression is low and the normalised variation between the highest and lowest points is only about 1.8. This suggests that *DtFLC* is not diurnally regulated.

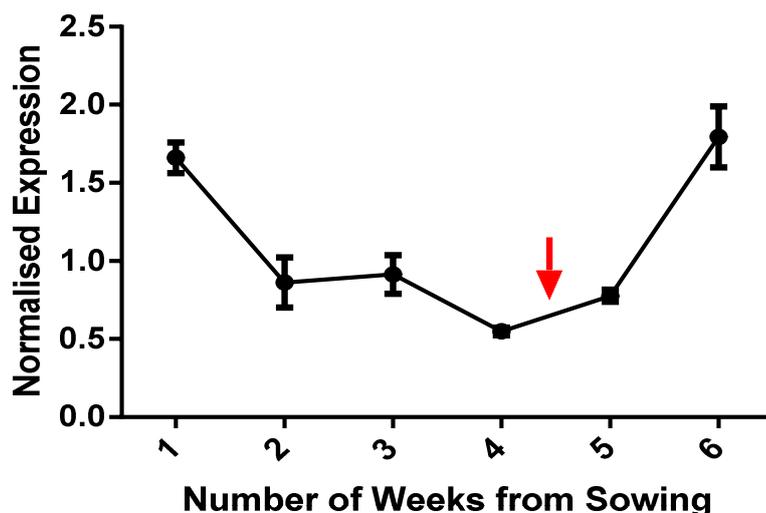


**Figure 3.21 – Diurnal Time Course of *DtFLC*.**

Expression of *DtFLC* over 52 hours is indicated to show the change in expression levels within 24 hours in response to the photoperiod. Lights were switched on at ZT0 (white bar) and off at ZT16 (black bar). Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars show the standard error of three technical replicates.

#### **3.3.3.4. Developmental time course**

Leaf material was collected at ZT15.5 at weekly intervals over the experimental period from sowing to two weeks post-initiation of bolting. Gene expression could then be analysed over development using these samples. The expression of *DtFLC* is fairly constant over development. After starting high it is then reduced and remains low until initiation where levels increase again (Figure 3.22). The difference between the highest and lowest expression is low indicating that there is only a small change in expression, but the lowering of *DtFLC* expression towards the time of bolting suggests there maybe be some influence over the transition from vegetative to reproductive states despite the low values.



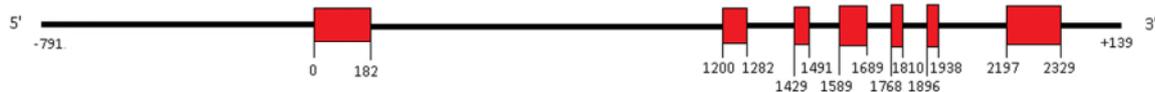
**Figure 3.22 – Expression of *DtFLC* across development.**

Leaf material was collected at ZT15.5 over six weeks from sowing to two weeks post-initiation of bolting. The red arrow indicates where the plants initiated bolting. Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars denote the standard error of three technical replicates.

### **3.3.4. SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)**

#### **3.3.4.1. Isolation of a rocket *SOC1* (*DtSOC1*)**

*DtSOC1* was successfully isolated using PCR and GeneRacer™ Kit (RLM-RACE) (Invitrogen). The full length CDS was found to be 642 nt in length with 7 exons (Figure 3.23, Appendix B.1.1.12). Excluding the 5' and 3' UTR, the *DtSOC1* gDNA is 2329 nt long (Appendix B.1.1.13). The full length CDS for *DtSOC1* when BLASTed was shown to have a 95% identity to *B. rapa* MADS-box *AGL20*. When translated, the protein sequence has a 95% identity to *B. juncea* MADS-box protein *AGL20/SOC1* and 93% identity to *A. thaliana* *SOC1*, with a MADS-box at the N-terminus followed by a K-box domain (Figure 3.24, Appendix B.1.1.14). These results add to findings by Kim et al. (2003) that *SOC1* is highly conserved across the *Brassicaceae* family with at least 90% homology to *A. thaliana* *SOC1*. Kim et al. (2003) also observed that the function of *SOC1* in respect to the flowering pathway was conserved.



**Figure 3.23 – Structure of *SOCI* gene CDS from *Diplotaxis tenuifolia*.**

Red bars represent exons. Black lines represent introns and UTR.

MADS-box protein AGL20/SOC1 [Brassica juncea]					
Sequence ID: <a href="#">gb AFM77896.1</a> Length: 213 Number of Matches: 1					
Range 1: 1 to 213 <a href="#">GenPept</a> <a href="#">Graphics</a> <span style="float: right;">▼ Next Match ▲ Previous Match</span>					
Score	Expect	Method	Identities	Positives	Gaps
412 bits(1059)	7e-144	Compositional matrix adjust.	203/213(95%)	207/213(97%)	0/213(0%)
Query 1	MVRGKTQMKRIENATSRQVTF	SKRRNGLLKKAFELSVLCDAEVSLIIFSA	RGKLYEFASS	60	
Sbjct 1	MVRGKTQMKRIENATSRQVTF	SKRRNGLLKKAFELSVLCDAEVSLIIFS	K KLYEFASS	60	
Query 61	NMQDTIDRYLRHTKDRVSTKPVSEENLQHLKHEAANMMKKIEQLEASKRRLGEGIGSCS	120			
Sbjct 61	NMQDTIDRYLRHTKDRVSTKPVSEENLQHLKHEAANMMKKIEQLEASKRRLGEGIGSCS	120			
Query 121	IEELQQIEIQLEKSVKCI	RARKTQVFKDQIEQLKQEKALAAENKTLAEKCGSHMIEVWS	180		
Sbjct 121	IEELQQIEIQLEKSVKCI	RARKTQVFKDQIEQLKQEKALAAENKTLAEKCGSHMIEVWS	180		
Query 181	NKNQESGRGDEESSPSSEVETQLFIGLPCSSRK	213			
Sbjct 181	NKNQESGRGDEESSPSSEVETQLFIGLPCSSRK	213			

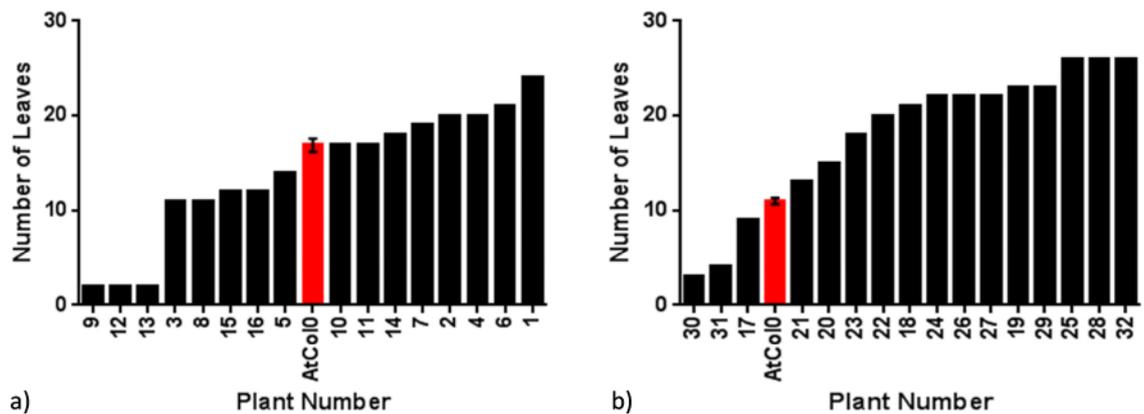
**Figure 3.24 – DtSOC1 protein BLASTP result.**

*DtSOC1* protein (Query) has been aligned to *Brassica juncea* SOC1 homologue (sbjct). Red line represents the MADS box domain and the blue line represents the K-box motif.

### 3.3.4.2. Complementation results

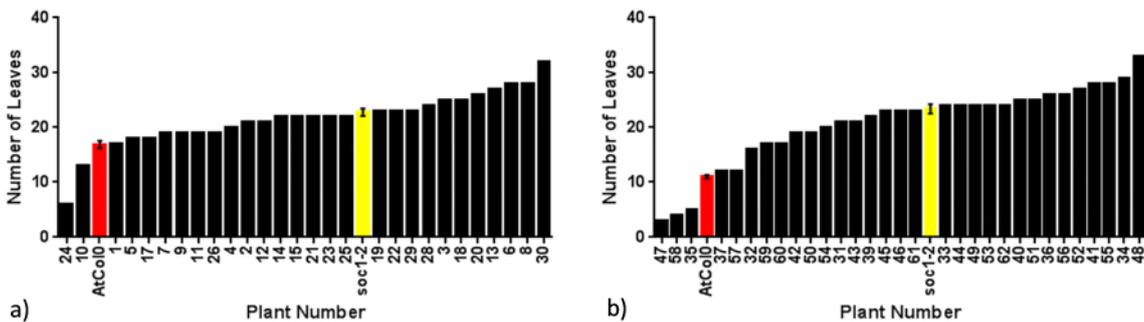
The isolated *DtSOC1* CDS was cloned into the destination vector pB2GW7 and transformed into *A. tumefaciens*. *AtCol0* wild type and *soc1-2* mutant plants were dipped into a solution of transformed *A. tumefaciens* cells and seed was collected from these plants for the T<sub>1</sub> generation. BASTA was used as the selection method and surviving seedlings were transplanted and scored for number of rosette leaves at flowering in LD. The T<sub>1</sub> screening was performed twice as many late flowering plants were recovered in the first screen which was unexpected. For *AtCol0:DtSOC1*, both T<sub>1</sub> screens showed a few plants which flowered after fewer days and with fewer leaves than the *AtCol0*, with most of the population flowering after more days and with more leaves than *AtCol0* wild type (Figure 3.25). However, the transgene PCR was able to amplify the transgene for all plants irrespective of flowering time, except *AtCol0:DtSOC1 5* (Figure 3.27). The results of the *soc1-2:DtSOC1* T<sub>1</sub> generation were very similar to that of *AtCol0:DtSOC1*. A few plants

flowered with fewer leaves than *AtCol0* wild type in both screens and a significant proportion flowered later, with more leaves than the *soc1-2* mutant (Figure 3.26). These data suggest that *DtSOC1* is able to complement the function of *AtSOC1*. The transgene PCR showed that the *DtSOC1* transgene was amplified in all plants tested across all flowering phenotypes. There was no transgene amplified in the *AtCol0* and *soc1-2* plants, or the negative control, so there was no contamination in the PCR or samples (Figure 3.27).



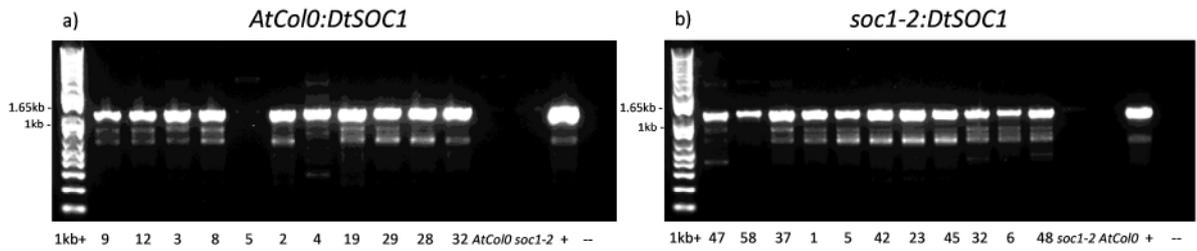
**Figure 3.25 – Flowering time data by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *AtCol0:DtSOC1* overexpression lines.**

a) First screen of T<sub>1</sub> generation plants of *AtCol0:DtSOC1*. b) Second screen of T<sub>1</sub> generation plants of *AtCol0:DtSOC1*. Red bar shows the *AtCol0* wild type flowering data (a- mean  $16.9 \pm 2.4$  n=12, b- mean  $11 \pm 1.1$  n=11). Error bars denote the standard error of the mean.



**Figure 3.26 – Flowering time data by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *soc1-2:DtSOC1* expression lines.**

a) First screen of T<sub>1</sub> generation plants of *soc1-2:DtSOC1*. b) Second screen of T<sub>1</sub> generation plants of *soc1-2:DtSOC1*. Red bar shows the *AtCol0* wild type flowering data (a- mean  $16.9 \pm 2.4$  n=12, b- mean  $11 \pm 1.1$  n=11) and yellow bar shows the *soc1-2* mutant flowering data (a- mean  $22.8 \pm 2.4$  n=12, b- mean  $23.4 \pm 2.8$  n=11). Error bars denote the standard error of the mean.

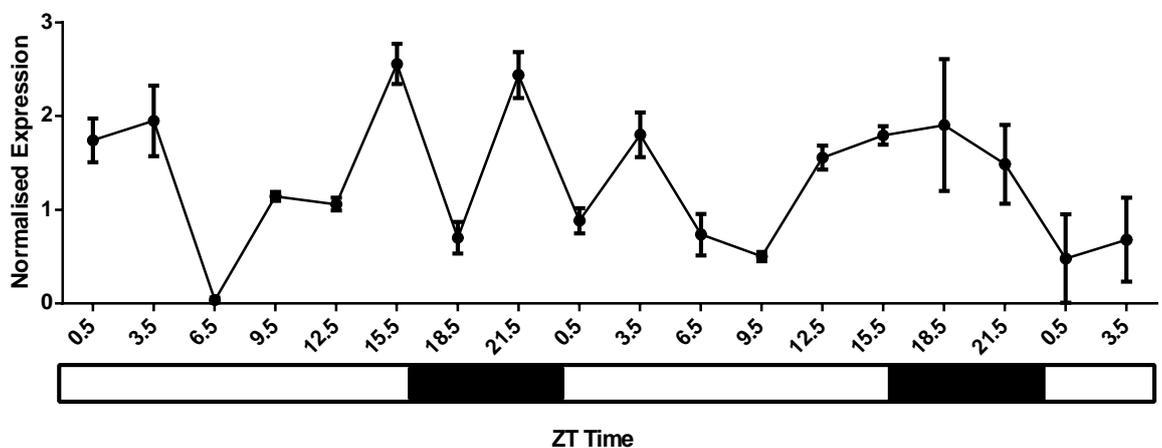


**Figure 3.27 – Transgene amplification in complementation line *AtCol0:DtSOC1* and *soc1-2:DtSOC1*.**

Gel pictures showing the amplification of the transgene by PCR for T<sub>1</sub> generation. a) Line *AtCol0:DtSOC1*, b) Line *soc1-2:DtSOC1*. Bands present at 1.2 kb show where transgene is present, absence of a band at 1.2 kb shows no transgene is present. 1 kb+ ladder used. Positive control (+) was generated using DNA extracted from a plasmid known to contain the *DtSOC1* transgene, the negative control (-) contained no DNA.

### 3.3.4.3. Diurnal time course experiment results

Analysis of the expression of *DtSOC1* over a 24-hour time period showed that there is only a small fluctuation in expression over 24 hours with a high expression at dawn followed by a dip in expression around ZT5-10, which then increases again overnight. However, the differences are small with the highest and lowest points only being 2.5 apart (Figure 3.28). There is variation between the same time points in each 24 hour cycle, so the results need to be repeated to get a clearer picture. However, the graph does suggest that *DtSOC1* may be diurnally regulated.



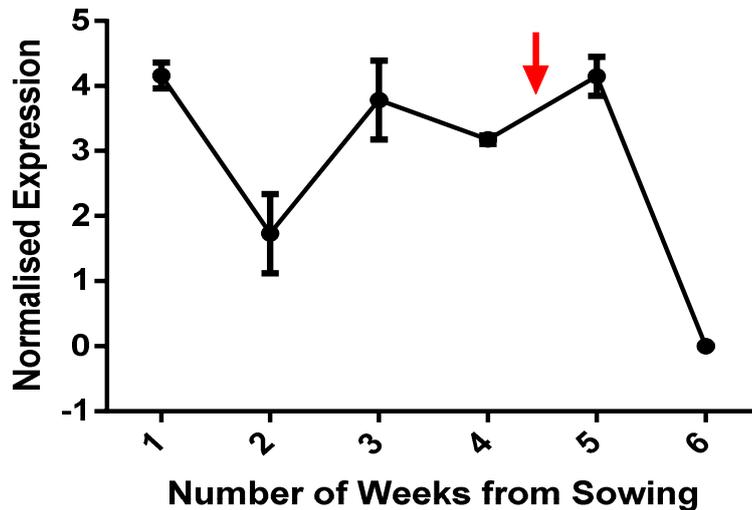
**Figure 3.28 – Diurnal Time Course of *DtSOC1*.**

Expression of *DtSOC1* over a LD diurnal time course of 52 hours. Lights were switched on at ZT0 (white bar) and off at ZT16 (black bar). Expression was normalised to the average expression of *Dta-tubulin*, *DtCACAS* and *DtTIP41*. Error bars show the standard error of three technical replicates.

### 3.3.4.4. Developmental time course

A time course was set up to examine the change in gene expression over the development of a rocket plant, from sowing until two weeks post initiation of

bolting. Leaf material was collected at weekly intervals and analysed using real time PCR. The expression of *DtSOC1* was quite variable over development. It starts quite high and falls a little in the early growth stages before increasing again as the plants mature. The levels don't change much as the transition to flowering occurs, but fall sharply once bolting has been initiated (Figure 3.29).



**Figure 3.29 – Expression of *DtSOC1* across development.**

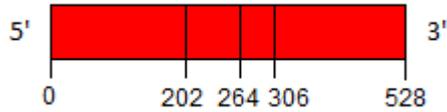
Leaf material was collected at ZT15.5 at weekly intervals. The red arrow indicates where the plants initiated bolting. Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars denote the standard error of three technical replicates.

### 3.3.5. FLOWERING LOCUS T (FT)

#### 3.3.5.1. Isolation of a rocket *FT* (*DtFT*)

The complete CDS for *DtFT* was successfully isolated from cDNA (Figure 3.30) using PCR and GeneRacer™ Kit (RLM-RACE) (Invitrogen). It was found to be 528 bp in length and made up of 4 exons (Appendix B.1.1.7). However, full length gDNA sequence for *DtFT* has not been isolated, so no intronic information is available. The full length CDS for *DtFT*, was BLASTed and revealed that the nucleotide sequence had a 99% identity to *A. thaliana FT* and 83% identity to *A. thaliana TSF*. When the alignment was done, it was found that the sequence of *DtFT* and *AtFT* only differed by 1 nucleotide, at position 387 (Figure 3.31). When translated, the *DtFT* protein was 100% homologous to the *A. thaliana* protein FT (Figure 3.32, Appendix B.1.1.8). This is because the nucleotide change does not affect the amino acid coded from the triplicate. The *DtFT* protein was also found to

have a 86% identity to *B. carinata* FT variant 9 and 82% identity to *A. thaliana* protein TSF.



**Figure 3.30 – Structure of *DtFT* gene CDS from *Diplotaxis tenuifolia*.**

Red bars represent exons. Exon boundaries are inferred from *AtFT* due to the homology of *AtFT* to *DtFT*.

Arabidopsis thaliana protein FLOWERING LOCUS T mRNA, complete cds					
Sequence ID: <a href="#">ref NM_105222.2</a> Length: 855 Number of Matches: 1					
Range 1: 83 to 610 <a href="#">GenBank</a> <a href="#">Graphics</a>					
Score	Expect	Identities	Gaps	Strand	
970 bits(525)	0.0	527/528(99%)	0/528(0%)	Plus/Plus	
Query	1	ATGTCTATAAAATATAAGAGACCCTCTTATAGTAAGCAGAGTTGTTGGAGACGTTCTTGAT			60
Sbjct	83	ATGTCTATAAAATATAAGAGACCCTCTTATAGTAAGCAGAGTTGTTGGAGACGTTCTTGAT			142
Query	61	CCGTTTAATAGATCAATCACTCTAAAGGTTACTTATGGCCAAAGAGAGGTGACTAATGGC			120
Sbjct	143	CCGTTTAATAGATCAATCACTCTAAAGGTTACTTATGGCCAAAGAGAGGTGACTAATGGC			202
Query	121	TTGGATCTAAGGCCTTCTCAGGTTCAAACAAGCCAAGAGTTGAGATTGGTGGAGAAGAC			180
Sbjct	203	TTGGATCTAAGGCCTTCTCAGGTTCAAACAAGCCAAGAGTTGAGATTGGTGGAGAAGAC			262
Query	181	CTCAGGAACTTCTATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCT			240
Sbjct	263	CTCAGGAACTTCTATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCT			322
Query	241	CACCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACAACCTGGAACAACC			300
Sbjct	323	CACCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACAACCTGGAACAACC			382
Query	301	TTTGCAATGAGATTGTGTGTTACGAAAATCCAAGTCCCACTGCAGGAATTCATCGTGTC			360
Sbjct	383	TTTGCAATGAGATTGTGTGTTACGAAAATCCAAGTCCCACTGCAGGAATTCATCGTGTC			442
Query	361	GTGTTTATATTGTTTCGACAGCTTGGCAGGCAAACAGTGTATGCACCAGGGTGGCGCCAG			420
Sbjct	443	GTGTTTATATTGTTTCGACAGCTTGGCAGGCAAACAGTGTATGCACCAGGGTGGCGCCAG			502
Query	421	AACTCAACACTCGCGAGTTTGCTGAGATCTACAATCTCGGCCTTCCCGTGGCCGCAGTT			480
Sbjct	503	AACTCAACACTCGCGAGTTTGCTGAGATCTACAATCTCGGCCTTCCCGTGGCCGCAGTT			562
Query	481	TTCTACAATTGTCAGAGGGAGAGTGGCTGCGGAGGAAGAAGACTTTAG			528
Sbjct	563	TTCTACAATTGTCAGAGGGAGAGTGGCTGCGGAGGAAGAAGACTTTAG			610

**Figure 3.31 – *DtFT* alignment result from blastn.**

BLAST searching *DtFT* as query revealed the highest homology was to *AtFT* (sbjct). Only 1 nucleotide difference was found between the two genes at nucleotide 387 (red box).

protein FLOWERING LOCUS T [Arabidopsis thaliana]						
Sequence ID: <a href="#">ref NP_176726.1</a> Length: 175 Number of Matches: 1						
Range 1: 1 to 175 <a href="#">GenPept</a> <a href="#">Graphics</a>			▼ Next Match ▲ Previous Match			
Score	Expect	Method	Identities	Positives	Gaps	
360 bits(925)	1e-124	Compositional matrix adjust.	175/175(100%)	175/175(100%)	0/175(0%)	
Query	1	MSINIRDPLIVSRVVGDVLDPPFNRSITLKVTYGQREVTNGLDLRPSQVQNKPRVEIGGED				60
Sbjct	1	MSINIRDPLIVSRVVGDVLDPPFNRSITLKVTYGQREVTNGLDLRPSQVQNKPRVEIGGED				60
Query	61	LRNFYTLVMVDPDVPSPSNPHLREYLHVLVTDIPATTGTTFGNEIVCYENPSPTAGIHRV				120
Sbjct	61	LRNFYTLVMVDPDVPSPSNPHLREYLHVLVTDIPATTGTTFGNEIVCYENPSPTAGIHRV				120
Query	121	VFILFRQLGRQTVYAPGWRQNFNTREFAEIYNLGLPVAAVFYNCQRESGCGGRRL				175
Sbjct	121	VFILFRQLGRQTVYAPGWRQNFNTREFAEIYNLGLPVAAVFYNCQRESGCGGRRL				175

**Figure 3.32 – DtFT protein BLASTP result.**

DtFT protein (Query) has been aligned to *Arabidopsis thaliana* FT homologue (sbjct). Green line represents the Phosphatidyl Ethanolamine-Binding Protein (PEBP) domain.

### 3.3.5.2. Isolation of a rocket *TSF* (*DtTSF*)

The complete coding sequencing of *DtTSF* was successfully isolated from cDNA using RLM-RACE, mRNA isolation and PCR. This gave a CDS of 528 nt (Figure 3.33, Appendix B.1.1.15). The four exon structures were assumed by aligning the CDS against *AtTSF*. The full length CDS for *DtTSF* was BLASTed and revealed that the nucleotide sequence had a 91% identity to predicted *B. rapa TSF-like* and 85% identity to *A. thaliana TSF*. It also had an 81% identity to the DtFT isolated gene. The nucleotide sequence was translated and BLASTed, revealing that the protein was 90% homologous to a predicted *B. rapa TSF-like* protein (Figure 3.34). It was also found to have an 81% identity to *A. thaliana* protein TSF and 80% identity to *A. thaliana* FT and DtFT proteins. The DtTSF protein contains the PEBP domain showing that is part of the same gene family as FT (Figure 3.34, Appendix B.1.1.16). The Blast programme consistently shows a high homology between DtTSF with both FT and TSF sequences, so an alignment was done to compare the similarities and differences of the isolated DtFT and DtTSF protein sequences against the AtFT, AtTSF and other genes in the PEBP family of AtTFL and AtBFT. Figure 3.35a compares the identity and divergence of each sequence and DtTSF has the greatest identity with AtTSF and the highest divergence with AtBFT. The phylogenetic tree in Figure 3.35b shows that the DtTSF is most closely related to AtTSF, followed by AtFT and DtFT. There are four key amino acids which differentiate function between PEBP family genes FT and TFL. These are aa85, and

segment B region amino acids aa134, aa138 and aa140. In AtFT and AtTSF, aa85 is a tyrosine but it is a histidine in AtTFL, in rocket it is a tyrosine in both DtFT and DtTSF (Figure 3.36). In other plant species where FT/TFL-like genes have been found, it has also been observed that there is a tyrosine at aa85 in the FT-like genes (Wickland and Hanzawa, 2015). A tyrosine is found at aa134 in both AtFT and AtTSF which is also observed in the FT inducer type species including both DtFT and DtTSF, however it is phenylalanine in AtTFL (Figure 3.36). A tryptophan is found at aa138 in AtFT and AtTSF which is also found in the FT inducer like genes including both DtFT and DtTSF, but not in AtTFL or repressor type FT genes. To have a repressor function, the FT/TFL like genes in other species had an amino acid change at either aa134 or aa 138, or both (Wickland and Hanzawa, 2015). This evidence suggests that both DtFT and DtTSF have the same inducer function as AtFT and AtTSF and not a repressor function like AtTFL. However, when aa140 is examined, in AtFT, DtFT and AtTSF the residue is a glutamine, but in DtTSF it is a proline and an aspartic acid in AtTFL (Figure 3.36). This difference may suggest a repressor like function of DtTSF, like that of Onion repressor FT-like gene AcFT4 where aa140 is a histidine (Lee *et al.*, 2013), but this isn't consistent across species as the sugar beet repressor FT-like gene BvFT1 is still a glutamine at aa140 (Pin *et al.*, 2010) as is the tobacco repressor FT-like gene NtFT1 (Wickland and Hanzawa, 2015).



**Figure 3.33 – Structure of *DtTSF* gene CDS from *Diplotaxis tenuifolia*.**  
Red bars represent exons, black lines represent UTR.

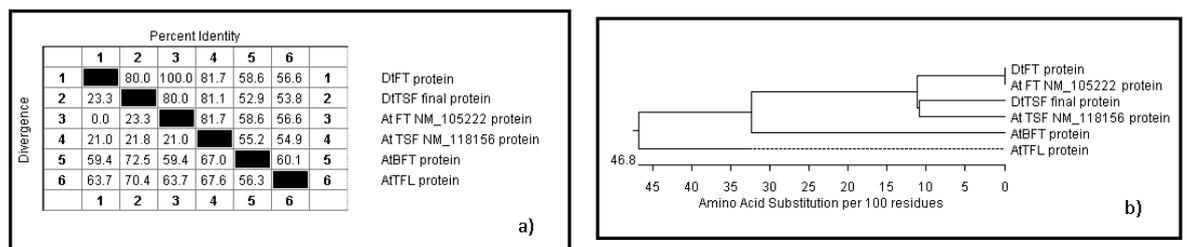
PREDICTED: protein TWIN SISTER of FT-like [Brassica rapa]  
Sequence ID: [reflXP\\_009106353.1](#) Length: 176 Number of Matches: 1

Range 1: 1 to 175 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
315 bits(806)	1e-106	Compositional matrix adjust.	158/175(90%)	168/175(96%)	0/175(0%)
Query 1		MSLSQRDPLVRLARVVGDVDPFMRLLISLRVITYGQRQVINGMDLRTSEVLNKPTVEIGGDD			60
Sbjct 1		MSLS RDPL++ RVVGDVDPFMRLLISL VTYGQRQ+ NGMDLR+S+VLNKPTVEIGGDD			60
Query 61		LRNFYTLVMVDPDVPSNP+LREYLHVLVTDIPATTGTAFGKELVVCYKNPCPTSGIHRV			120
Sbjct 61		LRNFYTLVMVDPDVPSNP+LREYLHVLVTDIPATTGT+FG ELV Y+NPCPTSGIHRV			120
Query 121		ALMLFRQLGRQTVYAPQWRPRFNTREFAENYNLGLPVAAVFFNCQRENGCGGRRT			175
Sbjct 121		AL+LFRQLGRQTVYAPQWR RFNTRERFAENYNLGLPVAAV+FNCQRENGCGGRRT			175

**Figure 3.34 – DtTSF protein BLASTP result.**

DtTSF protein (Query) has been aligned to *Brassica rapa* TSF-like homologue (sbjct). Green line represents the Phosphatidyl Ethanolamine-Binding Protein (PEBP) domain.



**Figure 3.35 – Sequence distances and phylogenetic tree comparing DtFT and DtTSF with AtFT, AtTSF, AtBFT and AtTFL.**

Protein sequences for FT-like genes compared using a) Sequence Distance (percent identity is the direct comparison of sequences and divergence is the comparison of sequence pairs with their phylogenetic relationship taken into account) and b) Phylogenetic Tree.

Consensus	PDVPSPSNP+LREYLHVLVTDI PATTGTTFGNEVVCYENPXPXTXGI HRVVFVXLF RQLGRQTVYAP- GWRQNFNT
At FT protein	PDVPSPSNP+LREYLHVLVTDI PATTGTTFGNEI VCYENPSPTAGI HRVVFV LFRQLGRQTVYAP- GWRQNFNT
DtFT protein	PDVPSPSNP+LREYLHVLVTDI PATTGTTFGNEI VCYENPSPTAGI HRVVFV LFRQLGRQTVYAP- GWRQNFNT
At TSF protein	PDVPSPSNP+HQREYLHVLVTDI PATTGNAFGNEVVCYESPRPPSGI HRI VLVLFRQLGRQTVYAP- GARQQFNT
DtTSF protein	PDVPSPSNP+LREYLHVLVTDI PATTGTAFGKELVVCYKNPCPTSGI HRVALMLFRQLGRQTVYAP- GWRPFFNT
AtTFL protein	PDVPGSPDPFLKEHLHVMVTNI PGTTDATFGKEVVSVELPRPSI GI HRFVFLFRQLGRQTVYAP- GWRQNFNT

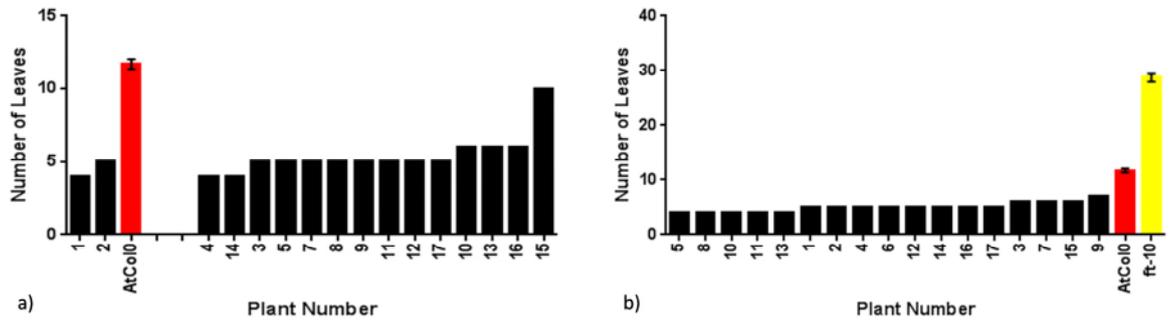
**Figure 3.36 – Protein alignments of AtTFL, AtFT, AtTSF with DtFT and DtTSF showing the key functional amino acids.**

Protein sequences for genes AtTFL, AtFT, AtTSF, DtFT and DtTSF aligned using MegAlign. Key amino acids at positions 85, 134, 138 and 140 in AtFT are highlighted in red boxes. In each case the only differences in amino acids are seen in the AtTFL sequence except for aa140 and 141 which also differs in DtTSF. Segment B which is conserved and corresponds to the external loop of FT is highlighted in the blue box.

### 3.3.5.3. Complementation of DtFT

The isolated *DtFT* CDS was transformed into *A. thaliana* mutant line *ft-10*. T<sub>1</sub> seed was selected using BASTA. Transformed plants were transplanted and flowering

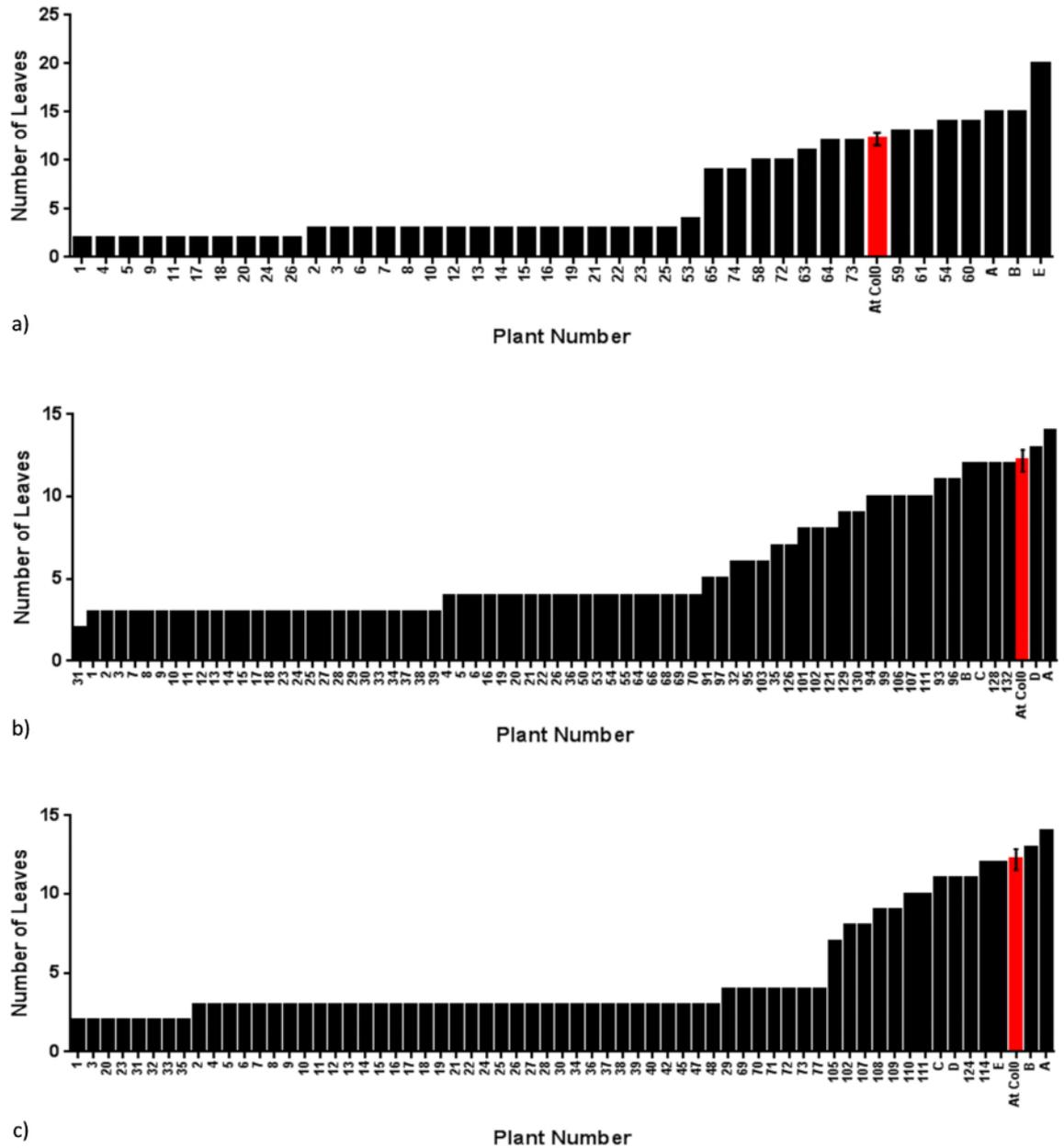
time in LD conditions recorded. The results for both *AtCol0:DtFT* (Figure 3.37a) and *ft-10:DtFT* (Figure 3.37b) transformations were very similar. All T<sub>1</sub> transformed plants flowered after fewer days and with fewer leaves than both *AtCol0* (red bar) and *ft-10* mutant (yellow bar) indicating that the *DtFT* transgene was able to promote flowering in either background. DNA was extracted from leaf material, but did not yield any results as the DNA and PCR from a number of samples was found to be contaminated.



**Figure 3.37 – Flowering time of T<sub>1</sub> generations of lines *AtCol0:DtFT* and *ft-10:DtFT* under LD.**

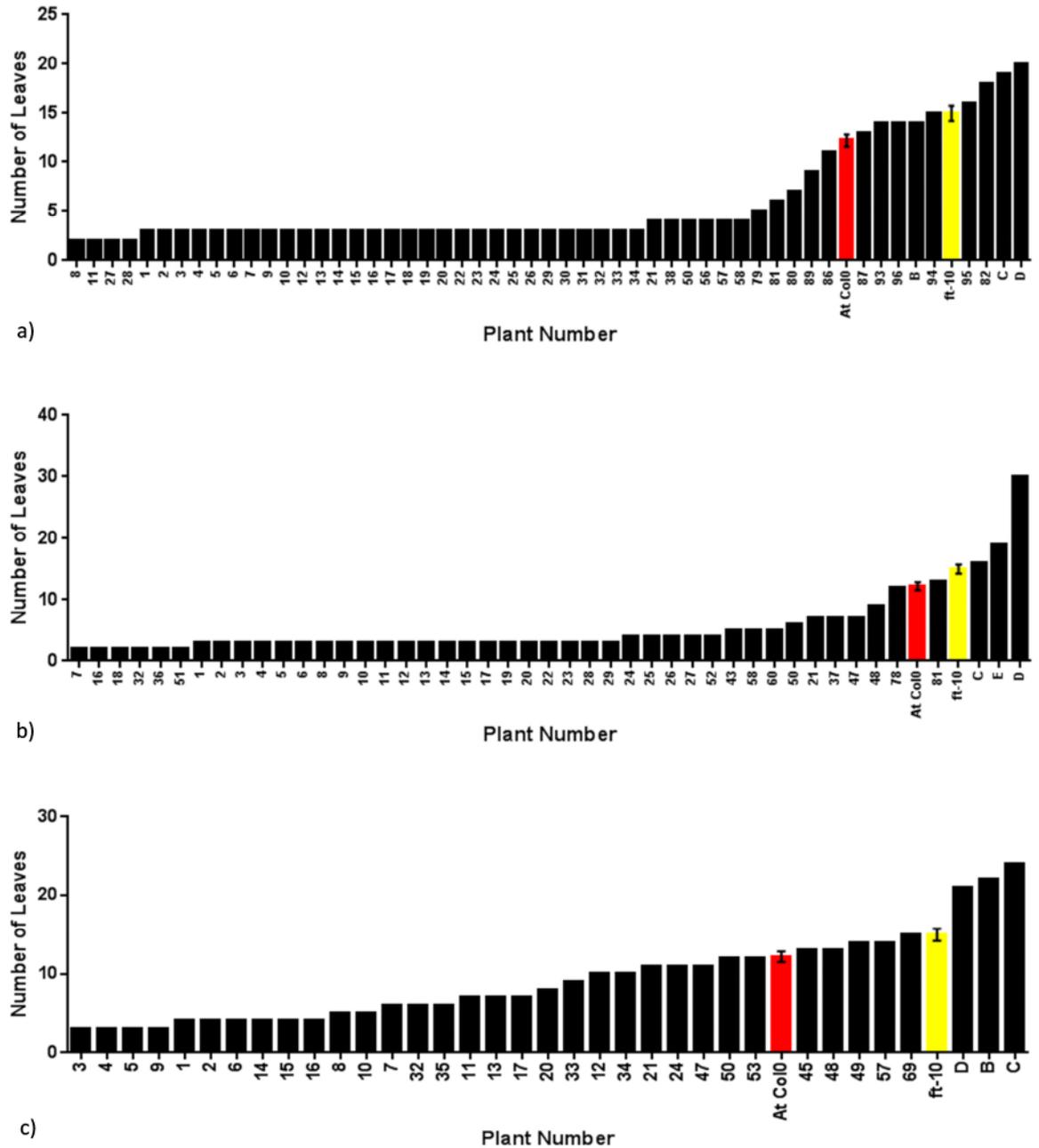
a) Flowering time data by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *AtCol0:DtFT* overexpression lines. *AtCol0:DtFT* lines 3-17 were sown at a later date without *AtCol0* wild type control seed. b) Bolting data by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *ft-10:DtFT* expression lines. Red bar shows mean number of leaves to flower for *AtCol0* wild type (mean  $11.7 \pm 1.5$  (n=20)). Yellow bar shows mean number of leaves to flower for *ft-10* mutant (mean  $28.7 \pm 2.2$  (n=9)). Error bars denote the standard error of the mean.

Seed from a number of lines of both *AtCol0:DtFT* and *ft-10:DtFT* was collected and sown out for the T<sub>2</sub> screen. All lines for T<sub>2</sub> generation of *AtCol0:DtFT* were found to be segregating with a 3:1 ratio of early:late flowering (Figure 3.38a, Figure 3.38b and Figure 3.38c) according to a Chi-squared goodness of fit test where the cut off for early flowering was six leaves. Lines *ft-10:DtFT 1* and *ft-10:DtFT 3* were also found to be segregating with a 3:1 early:late flowering ratio (Figure 3.39a and Figure 3.39b). However, line *ft-10:DtFT 5* was found to be significantly different from this ratio at  $p < 0.05$  (Figure 3.39c). As with the T<sub>1</sub> generation, DNA was extracted from leaf material but found to be contaminated, so PCR results were not trustworthy or conclusive as indicators of the presence or absence of the transgene. Due to this, two early flowering and one late flowering plants were chosen from T<sub>2</sub> lines *AtCol0:DtFT 4*, *ft-10:DtFT 1* and *ft-10:DtFT 5* to collect seed for screening in the T<sub>3</sub> generation.



**Figure 3.38 – Flowering time data for *AtCol0:DtFT* T<sub>2</sub> generation under LD.**

Flowering time data recorded by rosette leaf number when bolt was 1 cm. a) Flowering time data for line *AtCol0:DtFT 4*, b) Flowering time data for line *AtCol0:DtFT 8*, c) Flowering time data for line *AtCol0:DtFT 16*. Red bar shows mean number of leaves to flower for *AtCol0* wild type (mean  $12.2 \pm 1.5$  (n=5)). Error bars denote the standard error of the mean. Chi-squared test assessing goodness of fit for data to an expected 3:1 early:late bolting phenotype show that no lines were significantly different to the 3:1 ratio ( $p < 0.05$ ).

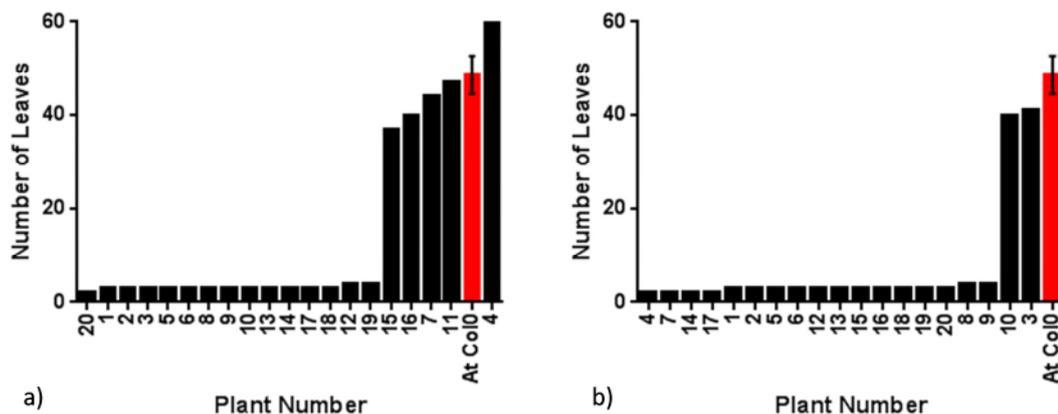


**Figure 3.39 – Flowering time data for *ft-10:DtFT* T<sub>2</sub> generation under LD.**

Flowering time data recorded by rosette leaf number when bolt was 1 cm. a) Flowering time data for line *ft-10:DtFT 1*, b) Flowering time data for line *ft-10:DtFT 3*, c) Flowering time data for line *ft-10:DtFT 5*. Red bar shows mean number of leaves to flower for *AtCol0* wild type (mean  $12.2 \pm 1.5$  (n=5)). Yellow bar shows mean number of leaves to flower for *ft-10* mutant (mean  $15 \pm 2$  (n=7)). Error bars denote the standard error of the mean. A Chi-squared test assessing goodness of fit for data to an expected 3:1 early:late bolting phenotype shows that lines *ft-10:DtFT 1* and *ft-10:DtFT 3* were not significantly different to the 3:1 ratio ( $p < 0.05$ ). Line *ft-10:DtFT 5* was significantly different to 3:1 ratio at  $p < 0.05$  if the cut off for early bolting is six leaves.

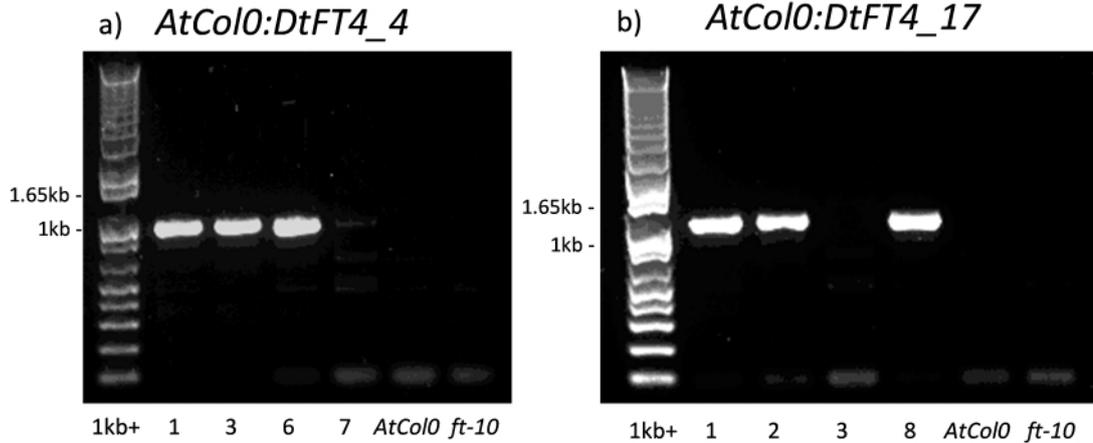
Complementation lines for *DtFT* T<sub>3</sub> seed were sown and grown under an 8 hour photoperiod (SD) at 22°C in a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd). This was done to slow the bolting of the early bolters so that a higher number of bigger leaves were available for DNA and RNA

extraction, and to reduce the expression of endogenous *AtFT* when the material was collected. Flowering time was scored as the number of rosette leaves when the bolt was 1 cm. Both *AtCol0:DtFT 4\_4* and *4\_17* were early flowerers at T<sub>2</sub> and this is also true in T<sub>3</sub> as the majority of plants are also early. However, the lines are still segregating as both contain a few late flowering plants in the population (Figure 3.40a and Figure 3.40b). DNA was extracted from several plants in each line for transgene PCR. The results from lines *AtCol0:DtFT 4\_4* and *4\_17* supported the flowering phenotypes observed as the early flowering plants did contain the transgene, and the late flowering plants did not. RNA was also extracted from leaf material of several plants in each line and real time PCR performed on cDNA synthesised from this RNA. Unfortunately, due to the small number of leaves on the plants, it was not possible to extract RNA from the same lines as the transgene PCR was performed on. The high expression of the *DtFT* transgene in early flowering plants *AtCol0:DtFT4\_4\_9* and *4\_4\_12* supports the early flowering phenotype, and the low level and absence of the *DtFT* transgene in the later flowering plants in lines *AtCol0:DtFT4\_4\_15* and *4\_4\_16* also suggests the activity of *DtFT* is affecting the plants ability and time to flower (Figure 3.42). The *AtCol0* wild type and *ft-10* mutant plants have no expression of *DtFT* so the *DtFT* real time primers (Appendix A.2.1.1) are not able to detect the endogenous *AtFT*, so this is not adding to the expression levels of *DtFT* found in the transformed lines (Figure 3.42).

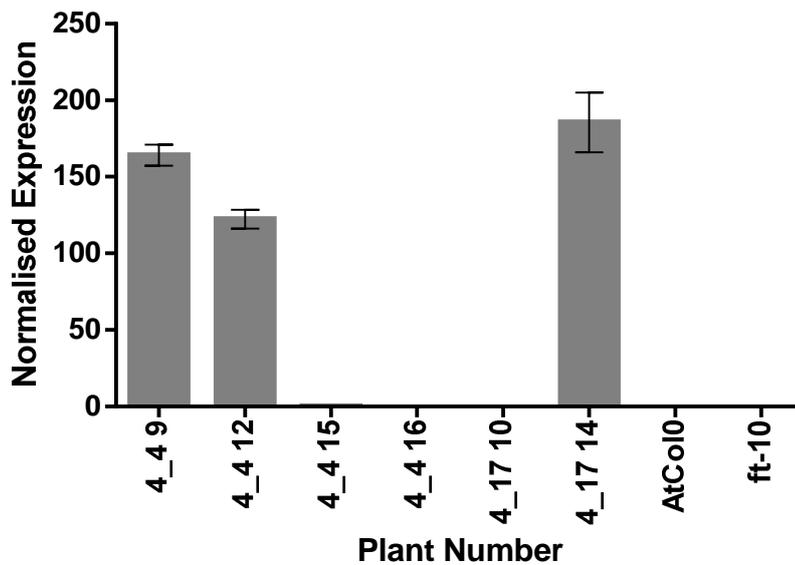


**Figure 3.40 – Bolting data for *AtCol0:DtFT4* T<sub>3</sub> generation under SD.**

Flowering time data recorded by rosette leaf number when the bolt was 1 cm. a) Line *AtCol0:DtFT 4\_4* flowering time data, b) Line *AtCol0:DtFT 4\_17* flowering time data. Black bars at 60 leaves did not flower before the end of the experiment so the actual number of leaves at bolting is unknown.



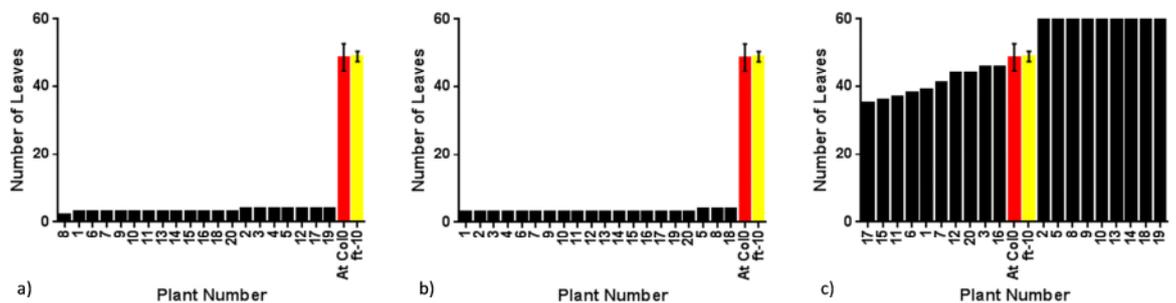
**Figure 3.41 – Transgene amplification in complementation line *AtCol0:DtFT 4*.**  
 Gel pictures showing the amplification of the transgene by PCR for three T<sub>3</sub> generation lines. a) Line *AtCol0:DtFT 4\_4*, b) Line *AtCol0:DtFT 4\_17*. Bands at 1 kb show where transgene is present. 1 kb+ ladder used.



**Figure 3.42 – Expression of *DtFT* transgene in *AtCol0:DtFT 4* T<sub>3</sub> generation under SD.**  
 Real time PCR amplification of *DtFT* transgene driven by 35S promotor is shown. Expression was normalised to the average expression of *Atβ-tubulin*, *AtActin2* and *AtTIP41*. Error bars are of standard error of three technical replicates.

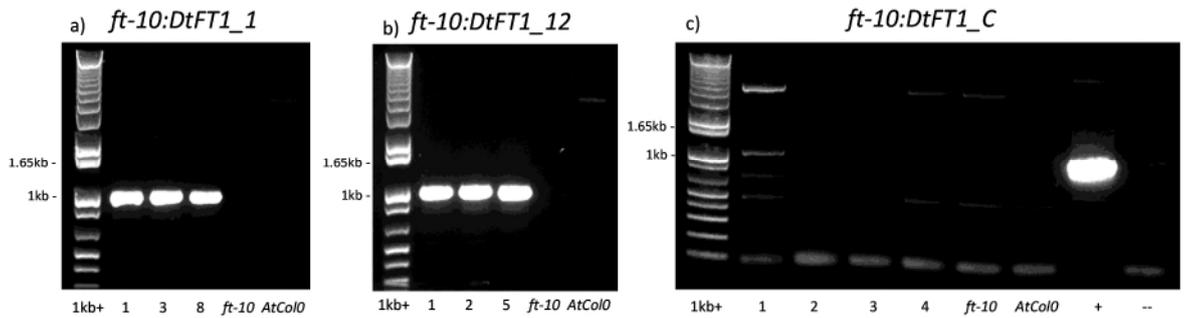
In the lines where *ft-10* was complemented with *DtFT*, the T<sub>3</sub> bolting data shows that the lines *ft-10:DtFT 1\_1* and *1\_12* are all transformed with the *DtFT* and that this has rescued the late flowering phenotype of the *ft-10* mutant. All T<sub>3</sub> plants are early flowering in both lines compared to *AtCol0* and *ft-10* under the 8-hour photoperiod, the lack of segregation could indicate that lines *ft-10:DtFT 1\_1* and *1\_12* may be

homozygous for the transgene (Figure 3.43a and Figure 3.43b). Figure 3.43c is the late line *ft-10:DtFT 1\_C* from the T<sub>2</sub> generation, and all plants in this line are still late, although not all are as late as *AtCol0* and *ft-10* mutant. All those plants which are later than the mutant and wild type did not flower by the end of the experiment. DNA extractions were done for several plants from each line and PCR used to detect the transgene. All plants tested showed that the transgene was present in early flowering lines and absent in late flowering lines, as expected (Figure 3.44). RNA was extracted from several plants from each line and real time PCR performed to investigate the expression levels of the *DtFT* transgene in the individuals. As with the *AtCol0:DtFT 4* experiments, it was not possible to extract RNA from the same plants tested in the transgene PCR. Figure 3.45 shows a high level of *DtFT* expression in all plants tested for lines *ft-10:DtFT 1\_1* and *1\_12* which supports the early flowering phenotype observed, that the transgene is causing this. Very low background levels of expression of *DtFT* was detected in *ft-10:DtFT 1\_C* plants *1\_C\_5* and *1\_C\_6* but this clearly was not sufficient to have any effect on flowering. There is no *DtFT* expression detected able in control *AtCol0* and *ft-10* mutant plants.



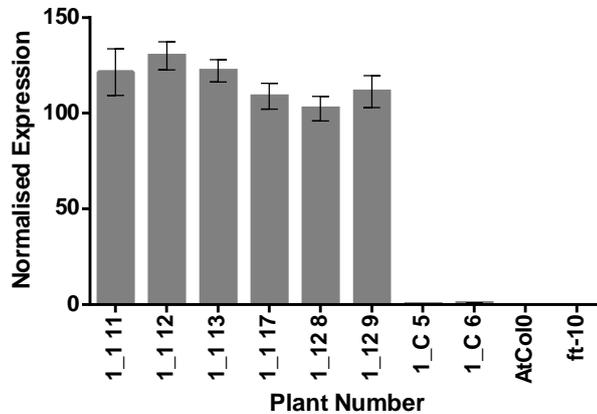
**Figure 3.43 – Flowering time data for *ft-10:DtFT 1* T<sub>3</sub> generation under SD.**

Flowering time data recorded by rosette leaf number when bolt was 1 cm. a) Line *ft-10:DtFT 1\_1* flowering time data, b) Line *ft-10:DtFT 1\_12* flowering time data, c) Line *ft-10:DtFT 1\_C* flowering time data. Red bar shows mean number of leaves at flowering for *AtCol0* wild type (mean  $49 \pm 4.3$  (n=8)). Yellow bar shows mean number of leaves at flowering for *ft-10* mutant (mean  $48.7 \pm 7$  (n=3)). Black bars at 60 leaves did not flower before the end of the experiment so the actual number of leaves at bolting is unknown. A one way ordinary ANOVA test found that there was a significant difference ( $p < 0.0001$ ) between the means of lines *ft-10:DtFT 1\_1* and *ft-10:DtFT 1\_12* compared to *ft-10:DtFT 1\_C*.



**Figure 3.44 – Transgene amplification in complementation line *ft-10:DtFT1*.**

Gel pictures showing the amplification of the transgene by PCR for three T<sub>3</sub> generation lines. a) Line *ft-10:DtFT 1\_1*, b) Line *ft-10:DtFT 1\_12*, c) Line *ft-10:DtFT 1\_C*. Bands at 1 kb show where transgene is present. 1 kb+ ladder used. Positive control (+) was generated using DNA extracted from a plasmid known to contain the FT transgene, negative control (-) is a water control.

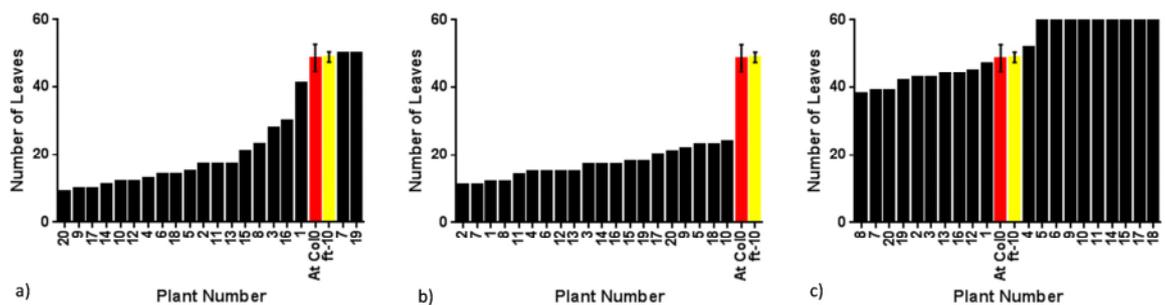


**Figure 3.45 – Expression of *DtFT* transgene in *ft-10:DtFT 1* T<sub>3</sub> generation under SD.**

Real time PCR amplification of *DtFT* transgene driven by 35S promoter. Expression was normalised to the average expression of *Atβ-tubulin*, *AtActin2* and *AtTIP41*. Error bars are of standard error of three technical replicates.

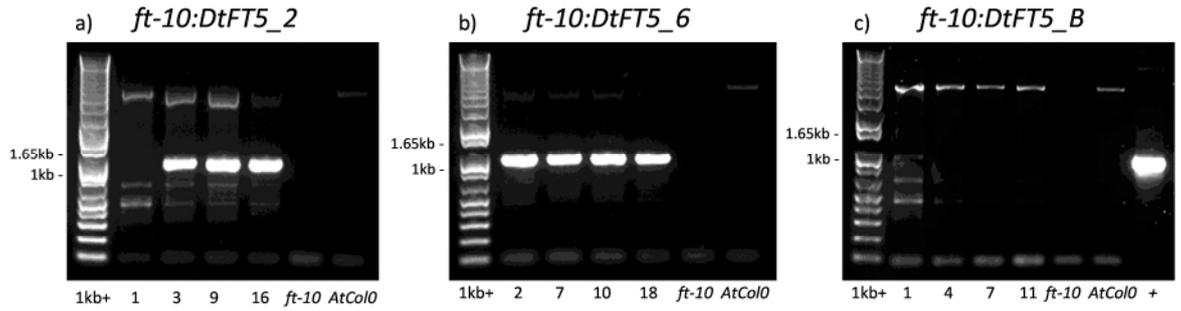
A screen of *ft-10:DtFT 5* T<sub>3</sub> generation was also carried out in the same way as for *ft-10:DtFT 1*. Flowering time data was collected from three lines *ft-10:DtFT 5\_2*, *5\_6* and *5\_B*. Figure 3.46a reveals that the early flowering phenotype of plants in line *ft-10:DtFT 5\_2* are not as extreme as *ft-10:DtFT 1\_1* and that the plants are still segregating for the transgene as two plants are later flowering than *ft-10* and those which are earlier range from 9 to 41 leaves at flowering. Plants in line *ft-10:DtFT 5\_6* (Figure 3.46b) are also not as early as those in line *ft-10:DtFT 1\_1* but all are earlier than *ft-10* with a small range of 11-24 leaves at flowering suggesting that this line might be homozygous for the transgene. Line *ft-10:DtFT 5\_B* was late at T<sub>2</sub> and is also late at T<sub>3</sub> (Figure 3.46c). Nine plants did not bolt before the end of the

experiment and the rest ranged from 38-52 leaves at flowering. DNA extractions were carried out for several plants from each line and the transgene detected using PCR. All plants tested showed that the transgene was present in early flowering plants and absent in late flowering plants (Figure 3.47). The expression levels of *DtFT* transgene in *ft-10:DtFT 5\_2* coincide with the flowering time data, where the higher the *DtFT* expression, the earlier the plant flowers, showing an obvious link between the two. *ft-10:DtFT 5\_6* is not so clear, but all plants tested have some *DtFT* expression and flower earlier than *ft-10* plants. *ft-10:DtFT 5\_B* plants *5\_B\_1*, *5\_B\_4*, *5\_B\_7* and *5\_B\_11* all show low background levels of *DtFT* expression, as does the *ft-10* plant, although they are late flowering, so it maybe that there is some slight contamination in the samples (Figure 3.48). Overall, the data from all the *ft-10* mutant complementation experiments, where the transgene is present, shows that *DtFT* is able to complement the function of *AtFT*.



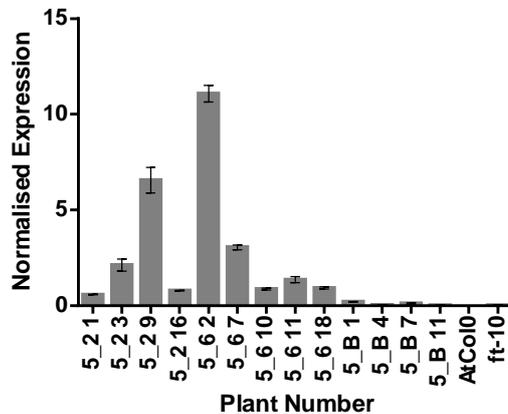
**Figure 3.46 – Flowering time data for *ft-10:DtFT 5* T<sub>3</sub> generation under SD.**

Flowering time data recorded by rosette leaf number when bolt was 1 cm. a) Line *ft-10:DtFT 5\_2* flowering time data, b) Line *ft-10:DtFT 5\_6* flowering time data, c) Line *ft-10:DtFT 5\_B* flowering time data. Red bar shows mean number of leaves at flowering for *AtCol0* wild type (mean  $49 \pm 4.3$  (n=8)). Yellow bar shows mean number of leaves at flowering for *ft-10* mutant (mean  $48.7 \pm 7$  (n=3)). Black bars at 60 leaves did not flower before the end of the experiment so the actual number of leaves at flowering is unknown. A one way ordinary ANOVA test found that there was a significant difference ( $p < 0.0001$ ) between the means of lines *ft-10:DtFT 5\_2* and *ft-10:DtFT 5\_6* compared to *ft-10:DtFT 5\_B*.



**Figure 3.47 – Transgene amplification in complementation line *ft-10:DtFT 5*.**

Gel pictures showing the amplification of the transgene by PCR for three  $T_3$  generation lines. a) Line *ft-10:DtFT 5\_2*, b) Line *ft-10:DtFT 5\_6*, c) Line *ft-10:DtFT 5\_B*. Bands at 1 kb show where transgene is present. 1 kb+ ladder used. Positive control (+) was generated using DNA extracted from a plasmid known to contain the FT transgene.



**Figure 3.48 – Expression of *DtFT* transgene in *ft-10:DtFT 5*  $T_3$  generation under SD.**

Real time PCR amplification of *DtFT* transgene driven by 35S promoter. Expression was normalised to the average expression of *At $\beta$ -tubulin*, *AtActin2* and *AtTIP41*. Error bars are of standard error of three technical replicates.

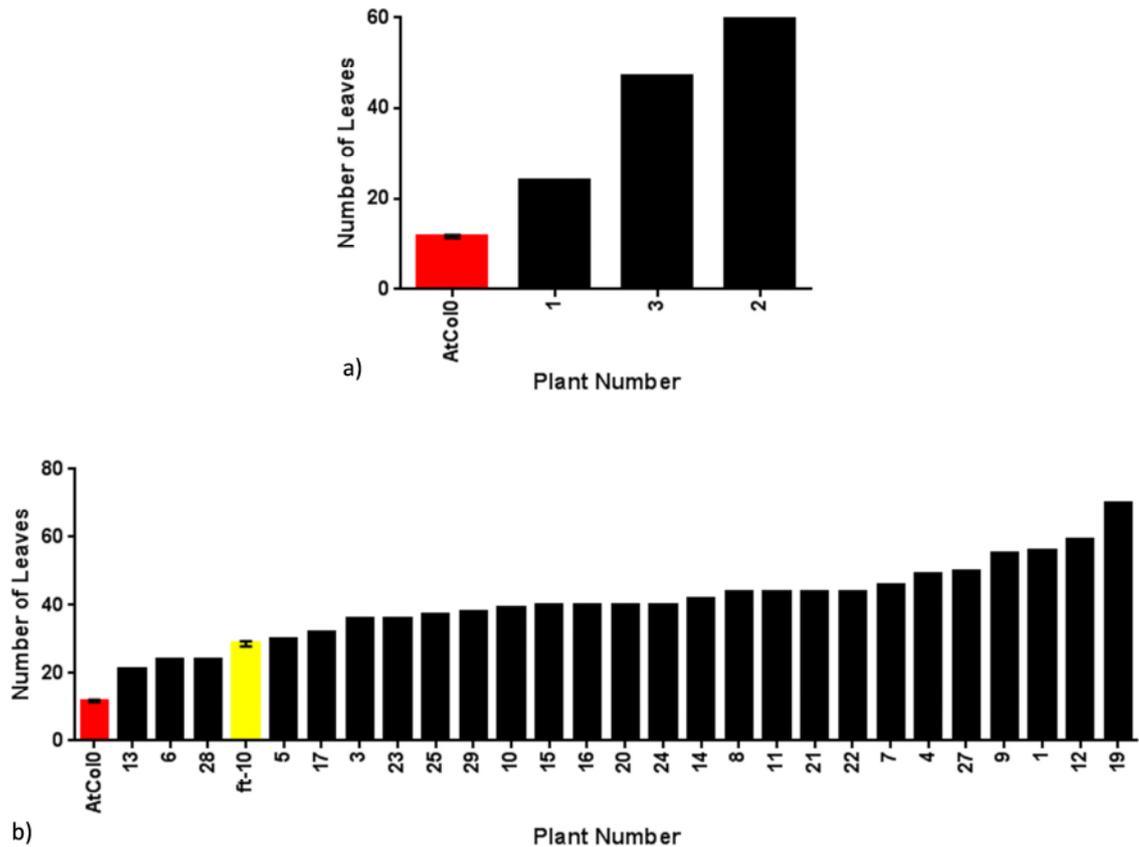
### 3.3.5.4. Complementation of *DtTSF*

The isolation of *DtTSF* CDS was not straightforward and two attempts were made. The first revealed the majority of the sequence but the 3' end was unconfirmed (*TSFfullLRcols*). The second attempt using an initial mRNA isolation step revealed the correct 3' end (*DtTSF final*). There was a one codon difference in the two sequences as shown by Figure 3.49. The first isolation product was transformed into *AtCol0* wild type and *ft-10* mutant. Unfortunately, there was not enough time to repeat the experiment with the *DtTSF final* product.  $T_1$  seed was selected using BASTA and the number of rosette leaves at 1 cm bolt was recorded.

Majority	ATGTCCTTAAAGTCAGAGAGATCCTCTTGTGCTTGCAAGAGTTGTGGGTGATGTTGTTGATCCTTTCATGAGATTGATCTC	
	10 20 30 40 50 60 70 80	
DtTSF final	ATGTCCTTAAAGTCAGAGAGATCCTCTTGTGCTTGCAAGAGTTGTGGGTGATGTTGTTGATCCTTTCATGAGATTGATCTC	80
TSFfullLRcols	ATGTCCTTAAAGTCAGAGAGATCCTCTTGTGCTTGCAAGAGTTGTGGGTGATGTTGTTGATCCTTTCATGAGATTGATCTC	80
Majority	TCTTAGAGTAACTTATGGCCAAAGACAGGTTATAAATGGAATGGATCTAAGGACTTCTGAAAGTTCTAAATAAACCAACGG	
	90 100 110 120 130 140 150 160	
DtTSF final	TCTTAGAGTAACTTATGGCCAAAGACAGGTTATAAATGGAATGGATCTAAGGACTTCTGAAAGTTCTAAATAAACCAACGG	160
TSFfullLRcols	TCTTAGAGTAACTTATGGCCAAAGACAGGTTATAAATGGAATGGATCTAAGGACTTCTGAAAGTTCTAAATAAACCAACGG	160
Majority	TCGAGATTGGAGGAGATGACCTCAGAAATTTCTACACTTTGGTTATGGTGGATCCAGATGTGCCAAGTCCAAAGCAACCCCT	
	170 180 190 200 210 220 230 240	
DtTSF final	TCGAGATTGGAGGAGATGACCTCAGAAATTTCTACACTTTGGTTATGGTGGATCCAGATGTGCCAAGTCCAAAGCAACCCCT	240
TSFfullLRcols	TCGAGATTGGAGGAGATGACCTCAGAAATTTCTACACTTTGGTTATGGTGGATCCAGATGTGCCAAGTCCAAAGCAACCCCT	240
Majority	CACCTCCGAGAAATATCTCCATTGGTTGGTGAATGATATACCTGCAACAACTGGAACAGCTTTTGGAAAAAGATTGGTGTG	
	250 260 270 280 290 300 310 320	
DtTSF final	CACCTCCGAGAAATATCTCCATTGGTTGGTGAATGATATACCTGCAACAACTGGAACAGCTTTTGGAAAAAGATTGGTGTG	320
TSFfullLRcols	CACCTCCGAGAAATATCTCCATTGGTTGGTGAATGATATACCTGCAACAACTGGAACAGCTTTTGGAAAAAGATTGGTGTG	320
Majority	CTATAAGAAATCCATGTCCCACTTCAGGAATTCACCGAGTTGCATTGATGTTGTTCCGACAACTCGGAAAGACAAACCGTTT	
	330 340 350 360 370 380 390 400	
DtTSF final	CTATAAGAAATCCATGTCCCACTTCAGGAATTCACCGAGTTGCATTGATGTTGTTCCGACAACTCGGAAAGACAAACCGTTT	400
TSFfullLRcols	CTATAAGAAATCCATGTCCCACTTCAGGAATTCACCGAGTTGCATTGATGTTGTTCCGACAACTCGGAAAGACAAACCGTTT	400
Majority	ATGCACCCCAGTGGCGCCACGCTTCAACACTCGTAGTTTCGCTGAGAACTACAATCTTGGTCTACCCGTTGCTGCCGTC	
	410 420 430 440 450 460 470 480	
DtTSF final	ATGCACCCCAGTGGCGCCACGCTTCAACACTCGTAGTTTCGCTGAGAACTACAATCTTGGTCTACCCGTTGCTGCCGTC	480
TSFfullLRcols	ATGCACCCCAGTGGCGCCACGCTTCAACACTCGTAGTTTCGCTGAGAACTACAATCTTGGTCTACCCGTTGCTGCCGTC	480
Majority	TTCTTCAACTGTCAAGAGAGAATGGCTGCGGAGGAAAGAAAGXXXXTAG	
	490 500 510 520	
DtTSF final	TTCTTCAACTGTCAAGAGAGAATGGCTGCGGAGGAAAGAAAGAACTAG	528
TSFfullLRcols	TTCTTCAACTGTCAAGAGAGAATGGCTGCGGAGGAAAGAAAGCTTAG	528

**Figure 3.49 – Alignment report comparing *DtTSF* cds isolated from *Dt* with TSFfullLRcols used in complementation studies.**

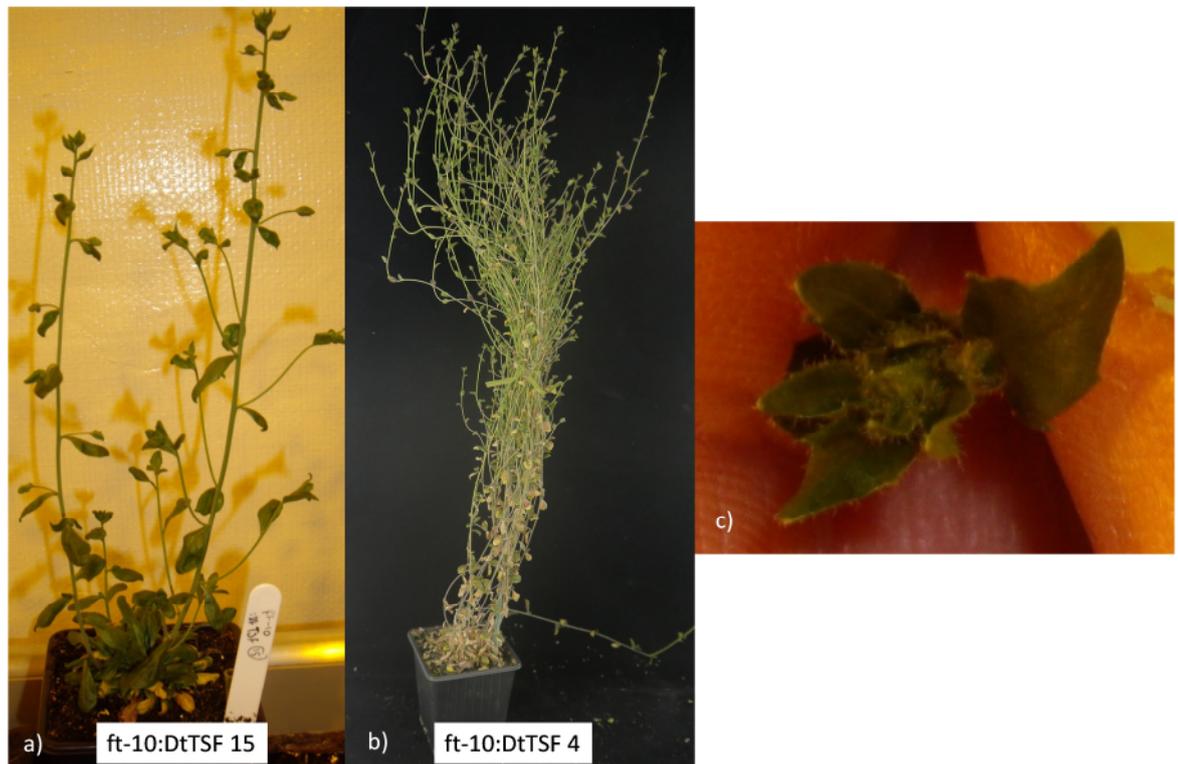
Complementation of *ft-10* mutant with *DtTSF* (TSFfullLRcols) began before the final sequence of *DtTSF* (DtTSF final) was isolated. Red box shows where the two sequences vary. The amino acid at the site of difference shown here is changed from Leucine in TSFfullLRcols to a Threonine in DtTSF final.



**Figure 3.50 – Flowering time of T<sub>1</sub> generation of *DtTSF* transformed into *AtCol0* wild type and *ft-10* mutant.**

a) Flowering time data by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *AtCol0:DtTSF* overexpression lines. b) Flowering time data by rosette leaf number at 1cm bolt for T<sub>1</sub> generation of *ft-10:DtTSF* expression lines. Red bar shows mean number of leaves to flower for *AtCol0* wild type (mean  $11.7 \pm 1.5$  (n=20)). Yellow bar shows mean number of leaves to flower for *ft-10* mutant (mean  $28.7 \pm 2.2$  (n=9)). Black bars at 60 leaves did not flower before the end of the experiment so the actual number of leaves at flowering is unknown. Error bars denote the standard error of the mean.

*AtCol0:DtTSF* transformed plants all bolted after more days and with more leaves than the *AtCol0* wild type (Figure 3.50a). In the T<sub>1</sub> transformation lines of *ft-10:DtTSF*, all plants also bolted after more days and with more leaves than *AtCol0* wild type and only three bolted with fewer leaves than the *ft-10* mutant plants but later in number of days. The rest bolted after more days and with more leaves than the *ft-10* mutant (Figure 3.50b). This suggests that the *DtTSF* gene is unable to complement the function of *AtFT*. The phenotype of *DtTSF* transformed plants showed many branches, and more inflorescence shoots which continued to grow, giving very tall plants (Figure 3.51a and Figure 3.51b). Many of the later bolting transformed lines produced leafy structures instead of flowers (Figure 3.51c) and although a few converted to flowers later on, very few lines were able to produce seed.

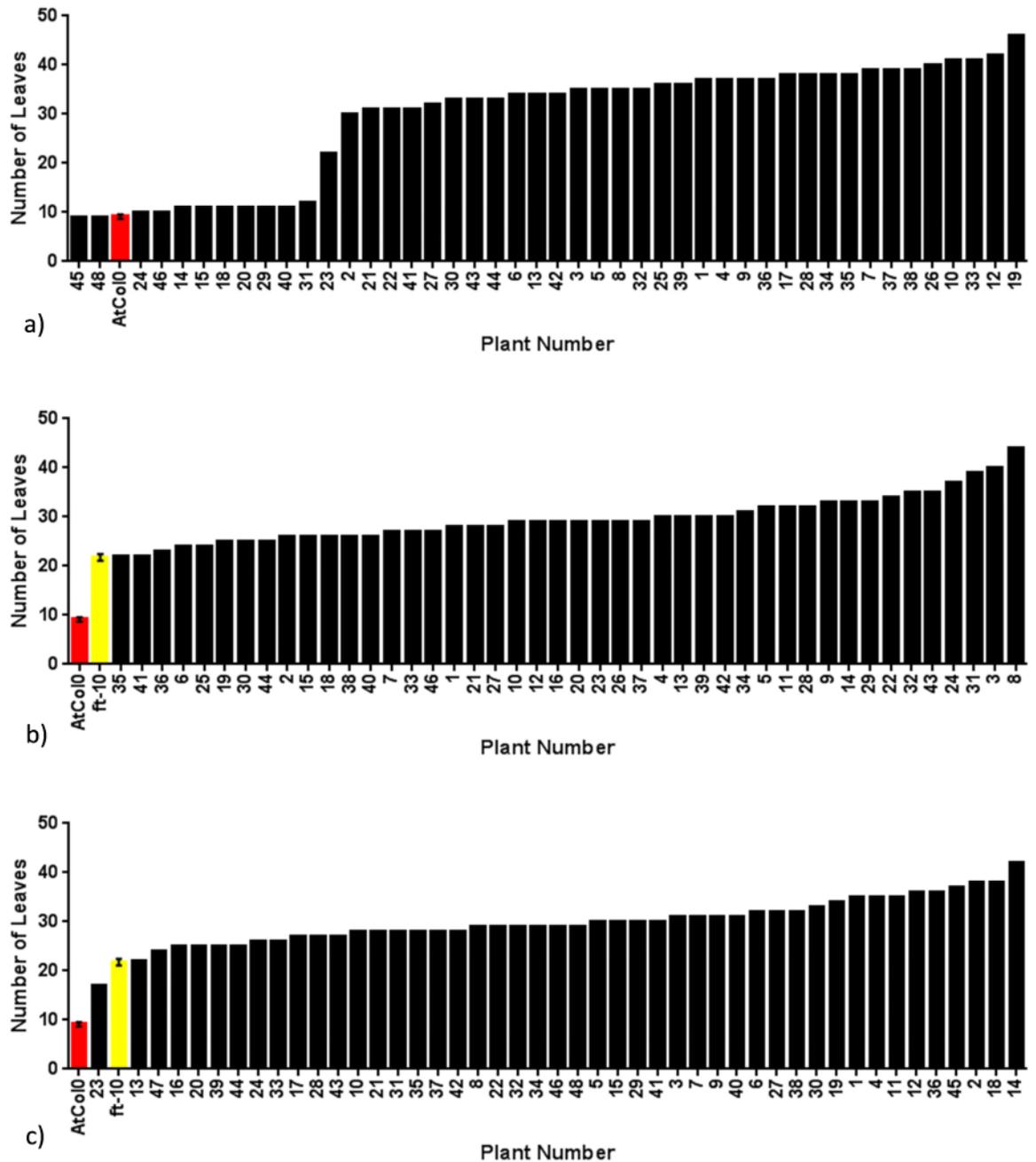


**Figure 3.51 – Inflorescence structures observed in *DtTsf* transformed plants.**

a) Photograph showing the bolted plant *ft-10:DtTsf 15* with highly branched inflorescences and no flowers. Photograph taken 62 days after sowing. b) Photograph showing the bolted plant *ft-10:DtTsf 4* with a large number of long inflorescences and no flowers. Photograph taken 109 days after sowing. c) A leafy type flower as observed on all *DtTsf* transformed plants. Photograph taken 62 days after sowing.

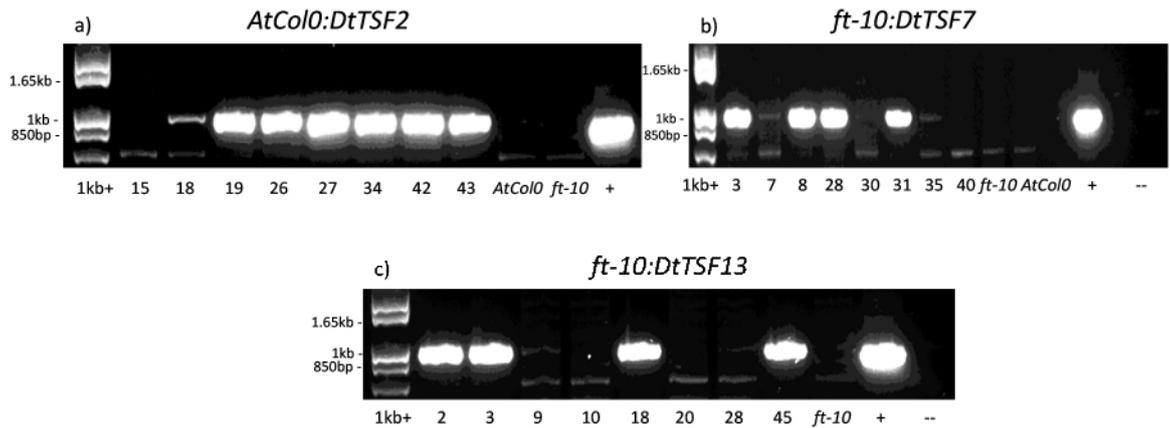
*AtCol0:DtTsf 2* bolted after the end of the experiment but was the only one which set seed and so was taken through to the next generation. Plants *ft-10:DtTsf 7* and *ft-10:DtTsf 13* both bolted within the experimental time and were two of a small number of plants which set seed. T<sub>2</sub> seed was sown out and screened for flowering time under a 16-hour photoperiod. Figure 3.52a shows the data for the T<sub>2</sub> generation of *AtCol0:DtTsf 2*. A Chi-squared goodness of fit test was applied and there was a significant difference ( $p < 0.05$ ) to the expected 3:1 early:late flowering ratio. This is unsurprising as the graph shows a large percentage of late flowering plants and no early flowering lines, although some flowered around the same time as WT. DNA was extracted from some of these lines and a PCR carried out to test whether the transgene was present in the samples. Figure 3.53a shows that the transgene was present in all late bolting lines, but no band in wild type bolting plant *AtCol0:DtTsf 2\_15* and a faint band in *AtCol0:DtTsf 2\_18*. This is unlikely to have been due to contamination, because there was no amplified band in the *AtCol0* wild type or *ft-10* mutant. The flowering time results for lines *ft-10:DtTsf 7* and *ft-10:DtTsf 13* were

very similar. All plants except one bolted with more leaves than the *ft-10* mutant (Figure 3.52b and Figure 3.52c). The Chi-squared goodness of fit test also showed a significant difference ( $P < 0.05$ ) to the expected 3:1 early:late bolting ratio where the cut off for the number of leaves at bolting for early flowering was 12. The transgene PCR results for *ft-10:DtTSF 7* and *ft-10:DtTSF 13* showed that those bolting around the time of the *ft-10* mutant did not contain the transgene and those which bolted much later than the *ft-10* mutant did contain the transgene (Figure 3.53b and Figure 3.53c). Together, this data suggests that *DtTSF* is unable to complement the function of *AtFT* and may even repress flowering.



**Figure 3.52 – Flowering time data by rosette leaf number at 1cm bolt for T<sub>2</sub> generation of DtTsf transformed lines.**

a) Flowering time data of line *AtCol0:DtTsf 2*, b) Flowering time data of line *ft-10:DtTsf 7*, c) Flowering time data of line *ft-10:DtTsf 13*. Red bar shows mean number of leaves to bolt for *AtCol0* wild type (mean  $21.8 \pm 2$  n=8). Yellow bar shows mean number of leaves to bolt for *ft-10* mutant (mean  $9 \pm 1.2$  n=8). Error bars denote the standard error of the mean. A Chi-squared test assessing goodness of fit for data to an expected 3:1 early:late bolting phenotype shows that all lines were significantly different to the 3:1 ratio ( $p < 0.05$ ) if the cut off for early bolting is 12 leaves.

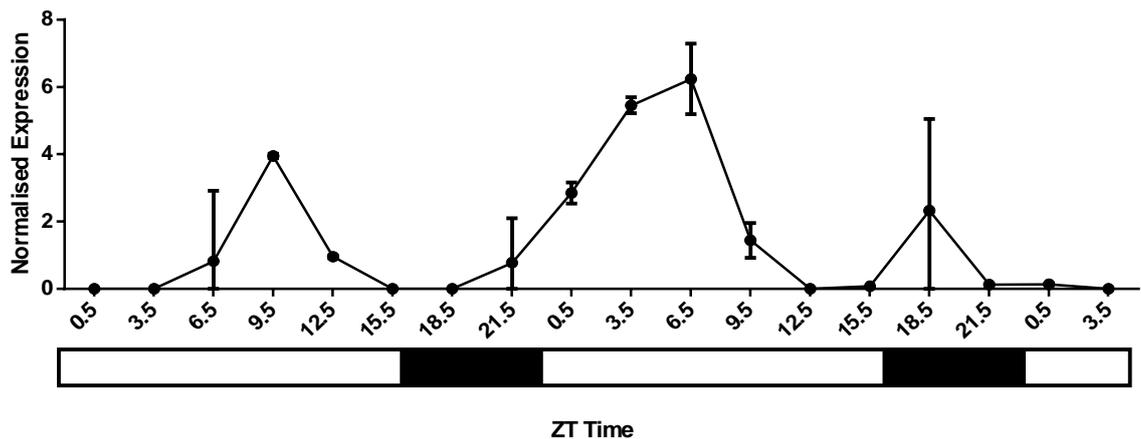


**Figure 3.53 – Transgene amplification in complementation line *AtCol0:DtTSF 2*, *ft-10:DtTSF 7* and *ft-10:DtTSF 13*.**

Gel pictures showing the amplification of the transgene by PCR for T<sub>2</sub> generation. a) Line *AtCol0:DtTSF 2*, b) Line *ft-10:DtTSF 7*, c) Line *ft-10:DtTSF 13*. Bands present at 1 kb show where transgene is present, absence of the white band at 1 kb shows no transgene is present. 1 kb+ ladder used. Positive control (+) was generated using DNA extracted from a plasmid known to contain the TSF transgene, negative control (-) is a water control.

### 3.3.5.5. Diurnal time course experiment

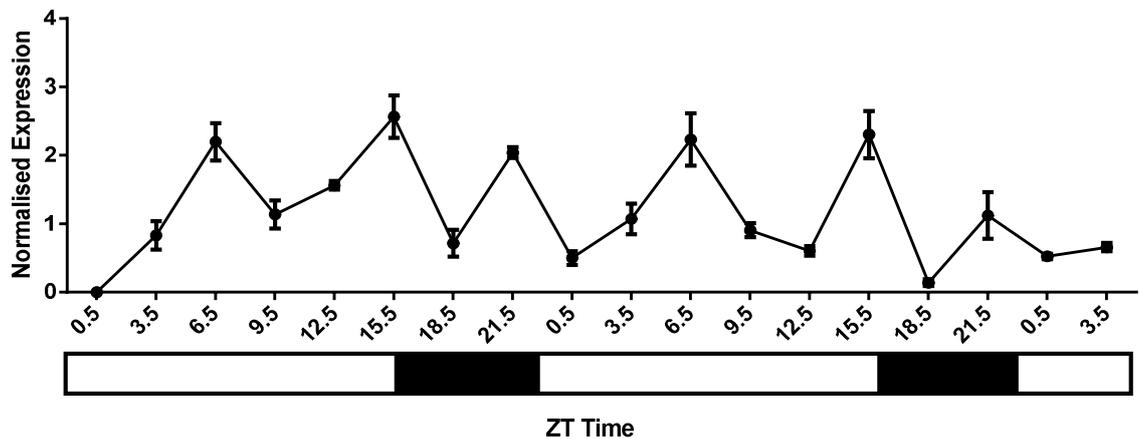
The expression of *DtFT* was measured using real time PCR over a 52 hour LD time course where plants were sampled every three hours. Figure 3.54 shows expression is highest around the middle of the light period and low just before, during and just after dark. This suggests that *DtFT* is diurnally regulated. However, some points have large error bars so it is difficult to see exactly what the pattern is.



**Figure 3.54 – Diurnal Time Course of *DtFT*.**

Expression of *DtFT* over a 52 hour time course in LD. Lights were switched on at ZT0 (white bar) and off at ZT16 (black bar). Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars show the standard error of three technical replicates.

The expression pattern of *DtTsf* was also investigated in this diurnal time course. Expression of *DtTsf* at dawn is low and then rises steadily to ZT 6.5. Expression then dips at ZT9-13 before rising towards the end of the light period. As night begins, expression falls until ZT18.5, then rises and falls again before dawn (Figure 3.55). The data therefore suggest that *DtTsf* is diurnally regulated.



**Figure 3.55 – Diurnal Time Course of *DtTsf*.**

Expression of *DtTsf* over a 52 hour time course in LD. Lights were switched on at ZT0 (white bar) and off at ZT16 (black bar). Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars show the standard error of three technical replicates.

### 3.3.5.6. Developmental time course

Leaf material was collected at weekly intervals at ZT15.5 and RNA extracted. This was used to analyse expression across the experiment from sowing until two weeks post-initiation of bolting. Expression levels of *DtFT* start very high, and then drop to levels which at week two, three, five and six which are barely detected in the PCR analysis. Expression levels do increase a little before initiation but drop again once this has occurred (Figure 3.56).



**Figure 3.56 – Expression of *DtFT* across development.**

Leaf material was collected at ZT15.5 at weekly intervals. The red arrow indicates where the plants initiated bolting. Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars denote the standard error of three technical replicates.

The change in expression of *DtTSF* was also measured over development at weekly intervals from sowing to two weeks post-bolting initiation. The expression of *DtTSF* seems to be linked to the transition to flowering as levels are low early in development before increasing rapidly until 1.5 weeks before initiation where levels sharply fall. Expression levels then stay low until the end of the experiment (Figure 3.57).



**Figure 3.57 – Expression of *DtTSF* across development.**

Leaf material was collected at ZT15.5. The red arrow indicates where the plants initiated bolting. Expression was normalised to the average expression of *Dta-tubulin*, *DtCACS* and *DtTIP41*. Error bars denote the standard error of three technical replicates.

### 3.4. Discussion

Five flowering time genes in Arabidopsis, *CO*, *GI*, *FLC*, *SOC1* and *FT* were selected as important genes in the flowering pathway in which to search for homologues in rocket. Sequences for each were isolated using PCR and RLM-RACE (Invitrogen) and used in complementation experiments of Arabidopsis mutants. Expression studies were also undertaken to investigate the change in gene expression over diurnal and developmental time courses. The genes *DtCO*, *DtGI*, *DtFLC* and *DtSOC1* were isolated with ease. Two *FT*-like sequences were isolated using the same methods followed by mRNA isolation and PCR to verify the differences observed between the two genes. These were named *DtFT* and *DtTSF* due to their high homology to each of the Arabidopsis *FT* genes *AtFT* and *AtTSF* respectively. The results of each of the flowering pathway gene homologues are discussed below.

#### 3.4.1. CONSTANS

Work has been carried out to try and isolate a *CO* homologue from rocket and test whether it has the same function as *AtCO* and *BrCO*. The isolation of *DtCO* using PCR and RLM-RACE (Invitrogen) proved successful with a full length CDS being isolated from cDNA and gDNA. *DtCO* was found to be highly homologous to other *Brassicaceae* *CO* orthologues (65-88% identity). The isolated *DtCO* CDS was then used in complementation studies and a diurnal and developmental time course.

The transformation of *DtCO* in *AtLer* wild type and *co-2* mutant backgrounds demonstrated that *DtCO* is able to restore flowering time to that of the wild type or earlier. This suggests that the *DtCO* gene functions as its orthologue in Arabidopsis does. This was shown in a similar experiment where a candidate *CO* gene in *B. napus* *BnCOa1*, was complemented into the *co-2* mutant (Robert *et al.*, 1998). Under both LD and SD conditions, the T<sub>3</sub> *BnCOa1* transgenic lines flowered with fewer leaves than the *co-2* mutant and a couple less than the *AtLer* WT. This difference was more pronounced however, under SD conditions. Therefore, the *DtCO* studies here should be extended to the T<sub>3</sub> generation and repeated using lines homozygous for the transgene in both LD and SD conditions.

Studies in Arabidopsis by Suarez-Lopez *et al.* (2001) found that *AtCO* expression was diurnally regulated. Under long day conditions, *AtCO* is high at dawn and drops quickly. It then rises again from ZT8 until it peaks at ZT16. Levels fall slightly at

ZT20 but are still relatively high and stay at this level until dawn. The group found that this oscillation in *AtCO* remains even when plants are exposed to constant light and thus this represents circadian regulation. The diurnal expression of *DtCO* was found to be similar to that seen of *AtCO* (Figure 3.9). It would be interesting to investigate whether this is the same in constant light and also whether the DtCO protein levels reflect the *DtCO* mRNA levels observed.

The developmental time course of *DtCO*, shows a rise in expression from one week after sowing until a peak at three weeks after sowing. Expression then falls by week four where bolting is initiated shortly after (Figure 3.10). Recent evidence in *Arabidopsis* suggests that the age dependent pathway through *miR172* is able to affect the expression of *CO* over development (Zhang *et al.*, 2015). In seedlings and young plants, levels of *miR172* are low and increase steadily with age. *miR172* is an inhibitor of the *TOE1* gene in the circadian clock. When *TOE1* is abundant early in development, *CO* expression is reduced. As the plants age, the level of *miR172* increases, which increases the inhibition of *TOE1*. This in turn releases the inhibition of *CO* by *TOE1* allowing the critical levels to be reached at the end of the long day promoting flowering (Zhang *et al.*, 2015). This age dependent pathway effect on *AtCO* could explain the developmental pattern of *CO* observed in rocket.

In conclusion, the data gathered in these experiments suggest that the isolated *DtCO* is the functional homologue of *AtCO* as it can successfully complement the function of *AtCO*, and is expressed in a similar way to *AtCO* in the diurnal and developmental time courses.

### 3.4.2. GIGANTEA

The *DtGI* gene was successfully isolated, cloned and sequenced from both cDNA and gDNA. It was found to encode a protein of 1179 amino acids, six larger than its homologue in *Arabidopsis*, and to be 89 and 92% homologous to both *Arabidopsis* and *B. rapa* genes respectively. This isn't surprising as both species and rocket are part of the *Brassicaceae* family. Attempts were made to complement the *Arabidopsis gi-3* mutant with the isolated *DtGI* gene but were unsuccessful. Recent work carried out by Xie *et al.* (2015) demonstrated that *BrGI* alleles are able to complement the *Arabidopsis gi-201* mutant. The late flowering phenotype of the *gi-201* mutant was rescued. The isolated *DtGI* gene was found to be 92% homologous to *BrGI* so it is expected that this homologue would be able to complement the *gi-*

201 mutant. A different mutant *gi-3* was used in our experiments but no transformed seed came through the BASTA selection. This could have been due to the size of the *DtGI* gene, decreasing the efficiency of transformation. The *A. tumefaciens* strain containing the *DtGI* transgene should be retested to confirm the presence of the gene in the vector. The mutant *gi-3* was poor at germinating, so a different knockout line could be used. The floral dip technique would be repeated in the new mutant line, perhaps with multiple dips as the plants continue to flower, to increase the number of transformed seed collected. The transformed lines would then be screened for flowering time to see if *DtGI* performs the same function as *AtGI*.

The diurnal regulation of *AtGI* was shown to oscillate over a long day with the lowest level at ZT0 and a large peak at ZT10 (Fowler *et al.*, 1999). This pattern has been observed for *DtGI* in the diurnal time course over long days demonstrating that it too is diurnally regulated. As *GI* is a known circadian clock gene, using the *DtGI* gene to complement *gi* mutants which are altered in circadian rhythm (such as *gi-1*, *gi-2* or *gi-201*) (Martin-Tryon *et al.*, 2007) could be performed to see if the *DtGI* transgene is able to restore the correct circadian rhythm. Gene-editing mutagenesis of *DtGI* could be attempted to create rocket plants impaired in the function of *GI*. This could further reveal the function of the *DtGI* homologue.

*AtGI* expression occurs throughout plant development as it is important for many growth and developmental changes due to its role in the circadian clock (Park *et al.*, 1999). It has been particularly implicated in the initiation of the circadian rhythm in seeds and fruit setting (Mishra and Panigrahi, 2015) which could explain the *DtGI* expression pattern seen in these experiments over development, where *DtGI* expression is constantly low throughout the majority of the plant's lifecycle, but with high peaks at week one seedling stage and week six which is post-bolting (Figure 3.14). However, the collection time for the developmental time course was at ZT15.5, whereas the diurnal time course showed that *DtGI* expression peaks at ZT10. Therefore the low expression of *DtGI* shown over development may be due to the low expression of the collection time and not a true picture of the expression change over the time course.

Further work needs to be undertaken to fully understand the nature and function of *GI* in rocket, particularly the complementation studies which are needed to confirm that the true rocket *GI* orthologue has been isolated. *AtGI* is a key photoperiodic

pathway and circadian clock gene which seems to share a high percentage of sequence with *DtGI* as well as its developmental and diurnal expression changes. Curtis *et al.* (2002) found that transforming an antisense fragment of *AtGI* into radish (*Raphanus sativus* – another *Brassicaceae* family member) caused a late bolting phenotype to be produced. This could be replicated in rocket to produce late bolting plants. These experiments could help to improve the understanding of *GI* and its role in controlling flowering in crop species.

### 3.4.3. FLOWERING LOCUS C

The aim of the study of *DtFLC* was to explore whether a gene homologous to *AtFLC* is present in rocket that is able to complement the Arabidopsis *FRI<sup>+</sup>flc3 null* mutant, and if so whether its expression over developmental and diurnal time courses was similar to that of other *FLC* homologues. The full length *DtFLC* gene was easily isolated from cDNA and gDNA using PCR and RACE-RLM (Invitrogen) and sequenced with gene specific primers. The *DtFLC* gene was found to be more homologous to *S. alba* and Brassica species *FLC* genes than to *AtFLC*.

*AtFLC* functions as a repressor of flowering and work done by Kim *et al.* (2007) investigated how candidate *FLC* genes in Chinese cabbage (*B. rapa* L. ssp. *Pekinensis*) behaved when complemented into Arabidopsis *flc* mutants. They found that all three *BrFLC* genes tested caused a late flowering phenotype showing that the candidate genes were able to complement the *flc* mutant and were repressors of flowering. A similar approach was taken to investigate the function of *DtFLC*. The *FRI<sup>+</sup> flc3 null* mutant in Arabidopsis as well as *AtCol0* wild type were transformed with the *DtFLC* transgene. The resultant seeds were selected using BASTA and assessed for flowering time. Plants displaying the expected late-flowering phenotype were selected and taken through to T<sub>2</sub> where flowering time was analysed and transgene expression investigated. The results revealed that *DtFLC* was able to complement the function of *AtFLC* in the mutant lines as a late flowering phenotype was observed in plants in T<sub>1</sub> and T<sub>2</sub> generations. Overexpression of *DtFLC* in *AtCol0* wild type was also found to cause late flowering, although early flowering lines were also observed in the T<sub>1</sub> (Figure 3.17a), suggesting there might be some silencing of the *AtFLC* gene in the *AtCol0* wild type. The transgene expression data showed that the late flowering phenotype is linked to the presence of the transgene in the plant.

*FLC* has been implicated in the regulation of the circadian clock through a small number of genes where the circadian period was shortened in plants grown at 27°C (Edwards *et al.*, 2006). This however, does not suggest that *FLC* itself is diurnally regulated or is part of the photoperiodic pathway (Srikanth and Schmid, 2011). The expression of *DtFLC* over a 24 hour time course suggests that *DtFLC* is also not diurnally regulated as there was very little change in expression over the light-dark cycle in long days (Figure 3.21).

As *FLC* is a floral repressor the expression of *FLC* during development needs to reduce in order for plants to be able to flower. Levels begin high to prevent precocious flowering but are reduced by vernalisation or autonomous pathways to a level where the inhibition of floral integrator genes *FT* and *SOC1* is released (Michaels and Amasino, 2001). The expression pattern of *DtFLC* reflects this, where levels are high initially but drop off towards the point at which initiation of bolting begins (Figure 3.22).

Overall, the experiments carried out on *DtFLC* have shown that the rocket homologue has been successfully isolated and retains the same function and expression patterns as *AtFLC* despite the homology between the two being only 84%. It has been encouraging that similar results have been seen for *DtFLC* as for Brassica species *FLC* genes as the homology between these is much higher.

#### **3.4.4. SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1**

The experiments undertaken to investigate the function and expression of *DtSOC1* demonstrate that a homologue of *SOC1* has been isolated from rocket. The protein contained the expected MADS-box and K-box domains with high homology to *B. juncea* (95%) and *A. thaliana* (93%). Complementation studies were carried out to the T<sub>1</sub> generation by transforming *DtSOC1* into both *AtCol0* wild type and *soc1-2* mutant plants. Many seedlings were selected as resistant to BASTA suggesting that they carried the transgene but the flowering time results showed only a few flowered with fewer leaves than *AtCol0* wild type plants (Figure 3.25 and Figure 3.26). In the *soc1-2:DtSOC1* population, only half the population flowered with fewer leaves than the *soc1-2* mutant. This suggests that the complementation may only be partial, so *DtSOC1* may not share all the functionality of *AtSOC1*.

The experiment investigating the change in expression of *DtSOC1* over time (Figure 3.29) showed a small increase in expression between week two and initiation of

bolting. This is different to what has been observed in Arabidopsis. *AtSOC1* expression increases steadily over development until flowering occurs (Lee *et al.*, 2000; Samach *et al.*, 2000). Under short day conditions, *AtSOC1* continues to be expressed after flowering, localised to the inflorescence. This specificity in location may explain why a drop in *DtSOC1* was observed after the initiation of bolting as leaf material was used to monitor the expression, not inflorescence material. It would be interesting to further investigate the location and timing of *DtSOC1* expression over development by sampling different tissues at weekly harvesting, and under both long and short day conditions.

The diurnal expression of *DtSOC1* was also investigated. The results showed a variable pattern where changes over the 24-hour period were small and a couple of time points did not have the same expression level in each of the two 24 hours' time periods examined. Therefore, it was concluded that *DtSOC1* could be diurnally regulated but the experiment will need repeating for verification. It would be interesting to repeat the experiment using short day conditions as well to see if the diurnal regulation is kept, or modified compared to long day conditions.

In conclusion, the work carried out to investigate the function and expression of *DtSOC1* has shown that the sequence isolated was highly homologous to *AtSOC1* and *BrSOC1* and some plants showed the restoration of the wild type flowering phenotype suggesting that *DtSOC1* is able to complement the *soc1-2* mutant and therefore the correct homologue. However, some plants did not show the expected phenotype which may be due to the site of integration of the transgene causing inhibition, or the overexpression causing co-suppression or interference in other key flowering time regulators.

#### **3.4.5. FLOWERING LOCUS T and FT-like genes**

The *FT*-like genes have key roles in flowering time in Arabidopsis. It was therefore important to isolate and investigate the function and expression of these in rocket, if a clearer understanding of flowering time was desired. Two genes, named *DtFT* and *DtTSF* were isolated using a GeneRacer™ Kit (RLM-RACE) (Invitrogen) and an mRNA isolation kit (Promega) followed by PCR, cloning and sequencing. The two CDS sequenced differed from each other and showed a high level of similarity to *AtFT* (99% homologous to *DtFT*) and *AtTSF* (83% homologous to *DtTSF*). It was found that the key amino acids which differentiated the function of FT and TFL all

matched FT in both DtFT and DtTSF, except for aa140 in DtTSF which was proline not a glutamine residue. As this is present in the segment B region of the FT gene, the exon 4 external loop, it may be enough to change the function from an inducer to a repressor type *FT*-like gene (Lee *et al.*, 2013; Wickland and Hanzawa, 2015). However, more evidence is needed to support this theory. Unfortunately, it was not possible to isolate intronic sequence for each gene so further differentiation of the genes using this information was not possible. The CDS for *DtFT* and *DtTSF* was used in overexpression studies in *AtCol0* wild type and complementation in *ft-10* mutant plants. Work done by Kobayashi *et al.* (1999) showed that the overexpression of *AtTSF* or *AtFT* in Arabidopsis gave a very early flowering phenotype. It was expected therefore that each candidate *FT-like* gene from rocket would also generate early flowering phenotypes. The overexpression and complementation results of *DtFT* did show the expected phenotype, but *DtTSF* transgenic plants showed a very late flowering phenotype with indeterminate growth (Figure 3.37-Figure 3.48 and Figure 3.50-Figure 3.53). This could be due to the silencing of *FT* and *TSF* genes in the plants by the *DtTSF* transgene, so that the *TFL* gene is more dominant and responsible for the phenotype (Baumann *et al.*, 2015). It could also be possible that the transgene of *DtTSF* is actually a *TFL*-like gene which would also explain the phenotype. However, the sequence similarity to AtTFL is only 70% and the key amino acids which confer inducer or repressor behaviour are all of the inducer type (Figure 3.36). Also, overexpression of *AtTFL* does not produce normal flowers (Hanano and Goto, 2011) but the aberrations are not the same as observed in the *DtTSF* complementation study (Figure 3.51). The difference in complementation results between *DtFT* and *DtTSF* are like those observed in sugar beet *BvFT1* (repressor) and *BvFT2* (inducer) (Pin *et al.*, 2010) and in onion *AcFT1* (inducer) and *AcFT4* (repressor) (Lee *et al.*, 2013). *DtFT* complemented the *ft-10* mutant and early flowering was observed, suggesting that *DtFT* is an inducer of flowering. *DtTSF* did not complement the *ft-10* mutant but very late flowering was observed suggesting that *DtTSF* is a repressor of flowering. This, together with the key amino acid residue data, suggests that two *FT*-like genes have been found in rocket, with opposing effects on flowering.

*FT* is a floral integrator gene which responds to daylength through the photoperiodic pathway. This means it is diurnally regulated as shown by Kobayashi *et al.* (1999). Experiments were undertaken in rocket using a 52 hour diurnal time course within

long day conditions. The diurnal time course of *DtFT* and *DtTSF* showed that *DtFT* was strongly regulated with a clear pattern of expression with a wide peak around ZT3-9 before falling towards the end of the light period (Figure 3.54). This is different to *AtFT* which peaks at the end of the long day and falls overnight (Song *et al.*, 2015). The *DtTSF* diurnal pattern is more difficult to see but is generally high during the light period before falling overnight, but seems to peak briefly in the middle of the night (Figure 3.55).

*FT* and *TSF* are both implicated in the promotion of flowering but their expression over development differs from each other. *AtFT* expression is low when the plants are young and increases as the plant ages with a peak seen before expression falls again (Kobayashi *et al.*, 1999). The *DtTSF* time course looks more like the *AtFT* developmental time course (Figure 3.57). *AtTSF* expression is high at early growth stages and falls as the plants age, expression then rises again (Kobayashi *et al.*, 1999). The *DtFT* developmental time course looks more like the *AtTSF* expression over development (Figure 3.56). Assessing these results suggests that the genes identified as *DtFT* and *DtTSF* may actually be mis-identified. However, the developmental time course experiment material was not collected at the optimal time for *DtFT* according to the diurnal expression pattern, so these results may be different if material was collected at ZT8.

Work by Yamaguchi *et al.* (2005) exploring further the role and function of *AtTSF* found that *TSF* and *FT* had a similar function but showed different spatial patterns of expression. *FT* was found to be predominantly expressed in leaves, whereas *TSF* was expressed in the hypocotyl. This distinctiveness may also be true in rocket, so it would be interesting to test for expression of each gene in different plant tissues. There was also a difference seen in expression across development, so different time points could be tested (Kobayashi *et al.*, 1999). This may help to confirm the identity of *DtFT* and *DtTSF*.

In conclusion, two *FT*-like genes (*DtFT* and *DtTSF*) have been isolated from rocket and tested for function and expression. The results were not always as expected and so it has not been possible to fully resolve which gene is the best candidate for *FT* and which for *TSF*. Further experiments are needed to continue to explore this, using a variety of tissues, time points and photoperiod lengths as well as further sequencing to isolate the gDNA sequences.

## **Chapter 4. Investigating the Photoperiodic and Vernalization Responses in Rocket**

### **4.1. Introduction**

The purpose of this chapter is to explore the effects of a changing environment on the growth of rocket and the corresponding expression changes of key flowering time pathway genes. This has not previously been investigated in rocket, but understanding these processes could lead to improved cultivation methods where the grower could have more control over harvest timing and bolting.

Two physiological traits are investigated and discussed. These are the vernalization and photoperiodic response. Both of these affect the plant's ability to flower and have a role in the flowering time pathway. Experiments were undertaken using controlled environment cabinets so that light, temperature and time could be manipulated, and plant material collected for gene expression analysis.

#### **4.1.1. Vernalization**

Vernalization is the attainment or acceleration of floral competence as a consequence of prolonged exposure to cold temperature. The cold treatment causes levels of *FLC* to reduce in the plant through epigenetic change in gene expression. When the plant is returned to a higher temperature, the change is stable enough to cause floral promotion rather than a return to high *FLC* levels (Boss *et al.*, 2004). Studies of *Arabidopsis* ecotypes showed that there are two groups with the split based on their vernalization responsiveness. Rapid cycling ecotypes will flower irrespective of vernalization, whereas winter annual ecotypes will flower late unless exposed to a period of vernalization (Nordborg and Bergelson, 1999; Wang, 2014). It was found that these flowering and vernalization responsive phenotypes all centre on the action of *FLC* (Song *et al.*, 2012). Plants which are responsive to vernalization tend to be those from seasonal climates where adaptation to extended periods of cold can be advantageous. The stage in the plant's lifecycle which is responsive to vernalization varies between species as does the critical length of treatment needed in order to fully satisfy the vernalization requirement. *Arabidopsis* is able to respond to vernalization both as seed and young plants (Nordborg and Bergelson, 1999; Strange *et al.*, 2011). *Brassica* species only respond as young plants which are beyond the

juvenile stage (Jung and Mueller, 2009). As rocket is closely related to both of these, it is not known at which stage of its lifecycle it is able to respond, and what the critical length might be, or even if it has a vernalization response at all. Therefore three treatment lengths were tested on both seed and young plants to investigate the vernalization requirements of rocket. The temperature of the vernalization treatment is also important as many species, such as *Arabidopsis*, vernalize at 4°C (Nordborg and Bergelson, 1999), whereas others such as *B. oleracea* var. *botrytis* have an optimum vernalization temperature of 10°C (Guo *et al.*, 2004). The Mediterranean plant olive (*Olea europaea*), was shown to vernalize at 10-13°C before flowering (Hackett and Hartmann, 1967), so two temperatures of 4°C and 10°C were investigated to test whether one would be more effective in satisfying the vernalization requirement and hastening floral initiation in rocket.

#### **4.1.2. Photoperiodic response**

The ability of a plant to measure and respond to the day length or photoperiod is critical to its growth and development, but particularly floral transition. Most plants fall into one of three main photoperiodic categories. The first are long day (LD) responsive plants which flower when the photoperiod is longer than a critical length. The second are short day (SD) responsive plants which flower when the photoperiod is shorter than a critical length. The third are day neutral plants which flower in spite of the photoperiod or night period length (Matsoukas *et al.*, 2012). In LD plants, flowering can occur quite quickly, but if LD plants are subjected to SD conditions, they will flower much more slowly (facultative) or not at all (obligate). In facultative LD plants such as *Arabidopsis* (Weigel, 1995), other pathways in the flowering network can compensate so that the expression of *FT* (and other floral integrator genes such as *SOC1*) is raised to the critical level for flowering. In obligate LD plants, floral initiation doesn't occur under SD conditions. During crop domestication and in breeding today, the role of the photoperiodic pathway and the genes within it has been really important (Bluemel *et al.*, 2015). Deviations in the 'normal' photoperiodic response of a species has enabled varieties of key crops, such as wheat and barley, to be grown all over the world from constant intermediate day conditions of the equator to the short day/long day seasonality of northern and southern latitudes (Srikanth and Schmid, 2011).

It is not currently known what the photoperiodic response of rocket is so experiments were carried out in this study to investigate it. Rocket plants were grown under three different photoperiods and the response measured in terms of number of days to initiate bolting. One experiment was carried out using low level incandescent lighting to provide the day extension of the 8h photoperiod in order to try and uncouple the effects of photoperiod length response from promotion of growth and phase transition resulting from high rates of photosynthesis. This is possible due to the difference in wavelengths of light emitted by incandescent bulbs (high far-red light) (Thimijan and Heins, 1983) and fluorescent bulbs (high red and blue light) (Runkle *et al.*, 2012).

## **4.2. Methods**

### **4.2.1. Rocket seed**

*D. tenuifolia* seed generation SSD4 (section 2.1.1) was used in the vernalization experiments at 4°C. A further two generations of single seed descent was carried out following these experiments giving rise to SSD6 seed (section 2.1.1). This was used in the vernalization experiments at 10°C and the photoperiodic experiments.

### **4.2.2. Vernalization experiments**

#### **4.2.2.1. Seed vernalization at 4°C**

Seed was sown onto damp paper towels and covered with aluminium foil and placed at 4°C for one, two or four weeks. At the end of the vernalization period, seeds were removed from the paper and put onto F<sub>2</sub>S soil in P40 trays at one seed per cell and placed in the glasshouse at 20°C with a 16 hour photoperiod. The germination date was recorded. Control seed to stay at ambient temperature was sown in the same way but was placed into the glasshouse at 20°C for one week by which time the seeds had germinated. These were then transferred to F<sub>2</sub>S soil in P40 trays at one seed per cell. At four weeks post germination, plants were transplanted into 5 inch pots of M2 soil to continue growing. Floral bud initiation and bolting was recorded. Plant material was collected once a week at ZT9. RNA was extracted from seeds pre-sowing and germinated seedlings at weeks one, two and four.

#### **4.2.2.2. Plant vernalization at 4°C**

For the plant vernalization experiment at 4°C, seed was sown directly to F<sub>2</sub>S soil in P40 cells and put in a controlled environment cabinet (Versatile plant growth MLR-351, SANYO Electric Co. Ltd) at 25°C to aid germination with 16 hour photoperiod for five days (Light level 104  $\mu\text{mol s}^{-2} \text{m}^{-1}$ ). The germination date was recorded and plants grown until three weeks old (four weeks after sowing) in a glasshouse at 20°C with 16 hour photoperiod. These were then transplanted to 5 inch pots of M2 soil. After one week, the plants were moved to 4°C with 16 hour photoperiod to vernalize for two, four or eight weeks. The plants were then returned to glasshouse conditions. The number of days to initiate and then bolt was recorded. Control plants were kept at 20°C with 16 hour photoperiod (ambient conditions) for the entirety of the experiment. Plant material was collected every week at ZT9 throughout the experiment. For weeks one to three, all material above soil level was sampled from up to six plants. From week four, newly expanded leaves were collected from five plants and pooled together.

#### **4.2.2.3. Seed vernalization at 10°C**

Seed was sown onto damp paper towels and covered with aluminium foil. This was then placed into a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 10°C in the dark. After two weeks, all seed had germinated, so was transplanted onto Arabidopsis mix soil in p24s and returned to 10°C but with very low light levels (36.2  $\mu\text{mol s}^{-2} \text{m}^{-1}$ ) as done in Nordborg and Bergelson (1999) and Lin *et al.* (2005; 1999)), so plants would stay healthy but not grow very much. Vernalization was performed at 10°C for two, four or six weeks in total and the plants transferred to a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) with 22°C and 16 hour photoperiod at the end of each treatment. At 5-6 true leaves, plants were transferred to 5 inch pots of M2 soil and grown until initiation and bolting. Control seed to stay at ambient temperature were kept continuously at 22°C in a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd). Germination of the control seedlings occurred after one week, so were transferred onto Arabidopsis mix soil in p24s and placed back into the cabinet at 22°C with 16 hour photoperiod until initiation and bolting occurred. As for the 4°C seed vernalization experiment, plant material was

harvested every week at ZT9 with RNA being extracted from seeds at pre-sowing and week one and germinated seedlings destructively sampled for week's two to six.

#### **4.2.2.4. Plant vernalization at 10°C**

Seed was sown directly into 5 inch pots of Arabidopsis mix soil and placed into a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C with a 16 hour photoperiod. At four weeks after sowing, plants had six to seven true leaves and were moved to a controlled environment cabinet (Panasonic) at 10°C with a 16 hour photoperiod to begin the vernalization. Treatments were for two weeks, four weeks and six weeks before returning the plants to ambient conditions in a controlled environment cabinet at 22°C with a 16 hour photoperiod until initiation and bolting. For plants grown as a control in ambient conditions, seeds were kept in a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C with a 16 hour photoperiod throughout the experiment. Initiation and bolting dates were recorded, and plant material collected throughout the experiment at weekly intervals at ZT9. This was done in the same way as 4°C plant vernalization experiment with week one and three destructively sampled of all material above soil level (from up to six plants) and from week four, newly expanded leaves were collected and pooled from three plants.

#### **4.2.3. Photoperiodic experiments**

Seeds were sown onto F<sub>2</sub>S soil in p24s and topped with vermiculite. These were placed into a controlled environment cabinet (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C with 16 hours light, 8 hours dark. The germination date was recorded and at 10 days after sowing, 10-12 plants were moved to each of three photoperiods (Table 4.1). In experiment 1, the first was under short day conditions with 8 hours light and 16 hours dark, the second was under intermediate day conditions with 12 hours light and 12 hours dark and the third was under long day conditions with 16 hours light and 8 hours dark. In experiment 2, three photoperiods were also used, but the difference between short, intermediate and long conditions with day extensions using incandescent bulbs (Bell Striplite, 30W)(Light levels 33  $\mu\text{mol}/\text{m}^2/\text{s}$ ) instead of the standard fluorescent tubes (Light levels 124  $\mu\text{mol}/\text{m}^2/\text{s}$ ) as used in experiment 1 (Table 4.1). The photoperiods were short day conditions with 8 hours light and 16 hours dark, intermediate day conditions with 8 hours light, 4 hours day extension and 12 hours dark and long day conditions with 8 hours light, 8 hours

day extension and 8 hours dark. All experiments were undertaken in controlled environment cabinets (Versatile plant growth MLR-352, Panasonic Co. Ltd) at 22°C. Plants were transplanted at three weeks from sowing into 5 inch pots of M2 soil and returned to the correct photoperiod. Floral initiation and bolting was recorded for each plant. Only floral initiation was recorded for experiment 2 as the day extension caused the stem to elongate before floral buds appeared, so an accurate bolting date couldn't be acquired.

**Table 4.1 – Light regimes for photoperiod experiments 1 and 2.**

The light and dark period timings with bulb type for each of the photoperiods tested in each experiment.

<b>Photoperiod length</b>	<b>Experiment 1</b>	<b>Experiment 2</b>
<b>Short day</b>	8 hours light (fluorescent bulbs)	8 hours light (fluorescent bulbs)
	16 hours dark	16 hours dark
<b>Intermediate day</b>	12 hours light (fluorescent bulbs)	8 hours light (fluorescent bulbs)
		4 hours day extension (incandescent bulbs)
	12 hours dark	12 hours dark
<b>Long day</b>	16 hours light (fluorescent bulbs)	8 hours light (fluorescent bulbs)
		8 hours day extension (incandescent bulbs)
	8 hours dark	8 hours dark

#### **4.2.4. Gene expression analysis using real time PCR**

RNA was extracted from leaf material of vernalization and developmental time course experiments using Z6 buffer method (Logemann *et al.*, 1987) as outlined in section 2.2.2. Synthesis of cDNA from 1 µg DNase treated RNA was done using a ThermoScript cDNA synthesis kit (Invitrogen) following the manufacturer's

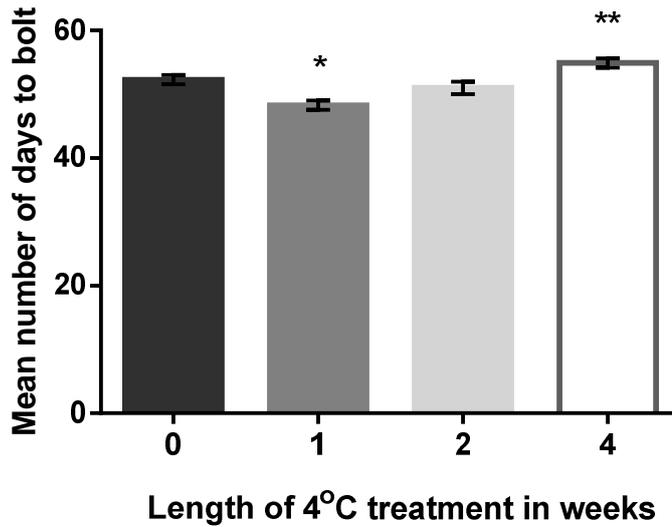
instructions. For the vernalization experiments, a mixture of RNA from samples where *DtFLC* expression is likely to be highest, were combined to synthesize cDNA to make the standard for the real time reactions. The standard was diluted 10 fold from  $10^0 - 10^{-4}$ . All real time PCRs carried out for vernalization experiments used *DtTIP41*, *DtCAC5* and *Dta-tubulin* as housekeeping genes for normalisation (Appendix A.2.1.3). The real time PCR conditions were as outlined in section 2.2.10.

### **4.3. Results**

#### **4.3.1. Vernalization experiments**

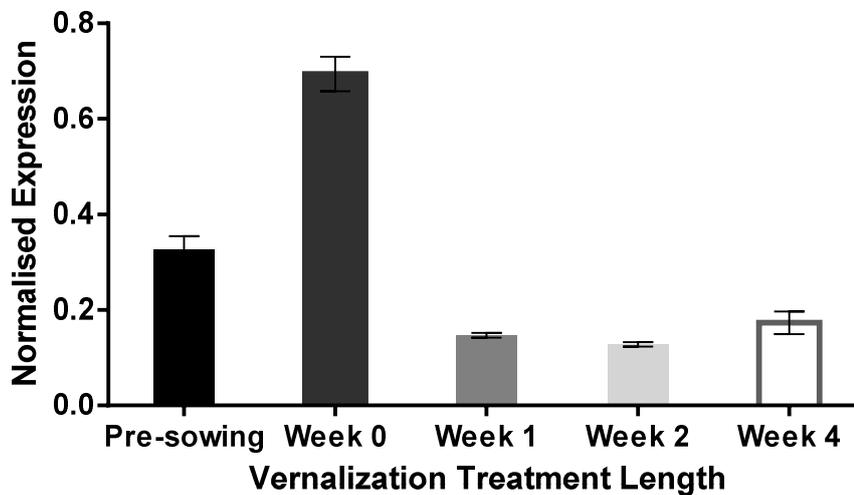
##### **4.3.1.1. Seed vernalization at 4°C**

Rocket seeds were subjected to a vernalization treatment of one, two or four weeks at 4°C before growing to maturity under ambient conditions. This was to assess whether the time to take to bolt would be quicker than that of control seeds which remained at ambient conditions for the duration of the experiment. Bolting data showed that there is no response of the seed to cold treatment of two weeks in length, but there is a significant difference for seed subjected to one ( $p < 0.0005$ ) and four ( $p < 0.0226$ ) weeks of cold (Figure 4.1). The average number of days to bolt in ambient temperature is very similar to that of a two week vernalization treatment, but with one week of cold, the average bolting time was quicker by four days and with four weeks vernalization treatment, the average number of days to bolt was delayed by 2.6 days. When *DtFLC* expression was analysed in the seed or seedlings (Figure 4.2), the results revealed that those subjected to one week of ambient temperature conditions had a higher level of expression than in pre-sown seeds and much higher expression than in the vernalization treatments. This suggests that it is the cold treatment that is affecting the *DtFLC* expression. However, there is little change between vernalization treatments in the level of *DtFLC* expression in the seedlings between weeks 1 to 4 indicating that the length of the vernalization treatment doesn't affect the *DtFLC* expression once the expression level has dropped initially.



**Figure 4.1 – Effect on bolting time of varying the length of 4°C cold treatment on *D. tenuifolia* seeds.**

The mean number of days to bolt of plants where the seed was subjected to cold treatment at 4°C for one, two or four weeks. Bars show mean ± SE (n=14 (0 weeks), n=15 (1 week and 2 weeks), n=10 (4 weeks)). Error bars denote the standard error of the mean. Student’s *t*-test was used to compare the number of days to bolt of each vernalization treatment (one, two and four weeks) against the ambient conditions (0 weeks). \*Statistical significance of  $p < 0.0005$  \*\* Statistical significance of  $p < 0.0226$ .

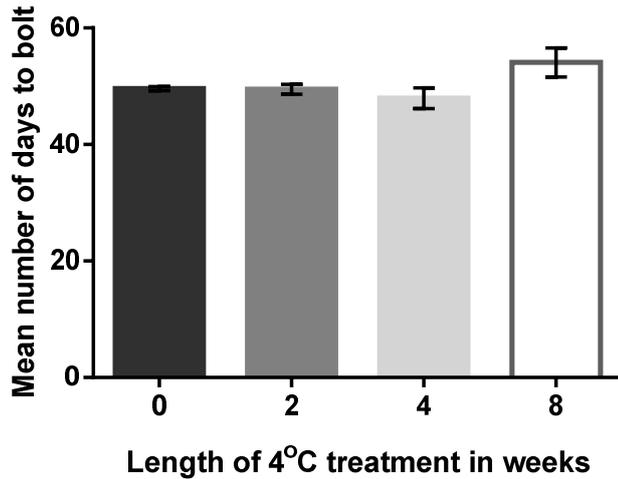


**Figure 4.2 – Effect of the different vernalization treatments at 4°C on expression of *DtFLC* in *D. tenuifolia* seed.**

Expression of *DtFLC* where seed was subjected to vernalization at 4°C for one, two or four weeks. Error bars denote the standard error of three technical replicates.

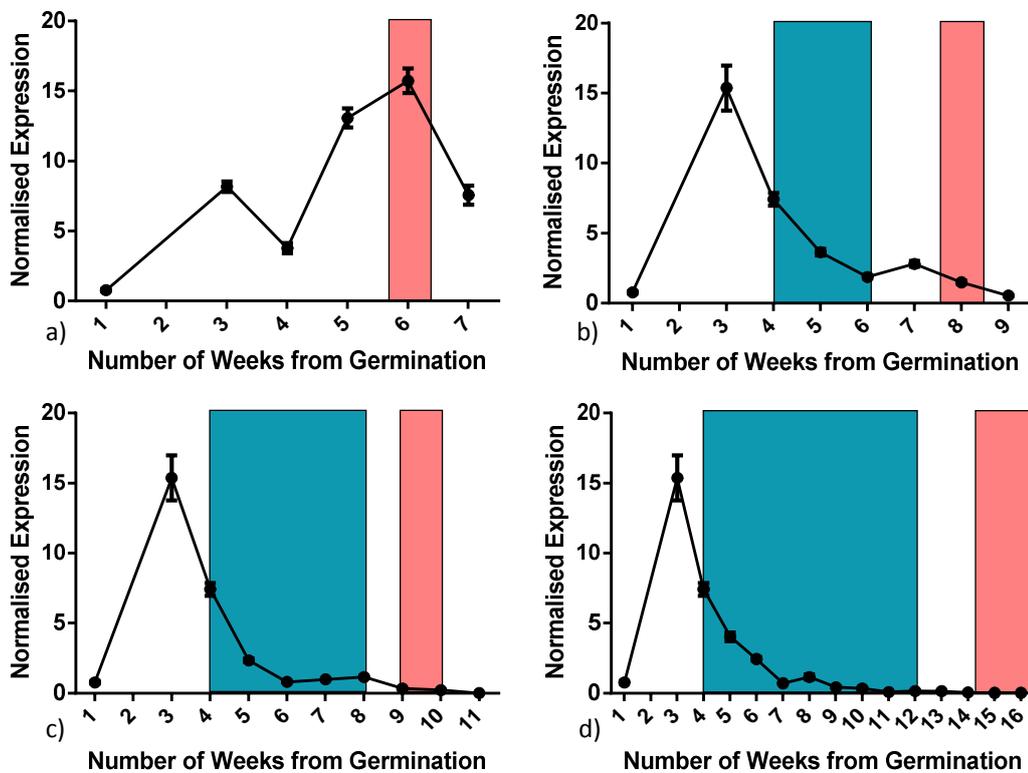
#### 4.3.1.2. Plant vernalization at 4°C

The effect of different vernalization treatment lengths at the temperature of 4°C was investigated for rocket plants. Seeds were sown and at four weeks after sowing, were placed at 4°C for two, four and eight weeks. A subset of plants remained at constant ambient temperatures throughout the experiment. There was no significant difference in bolting time between the plants kept at ambient temperatures to those subject to the different vernalization treatments (Figure 4.3). The plants which had been subjected to eight weeks of vernalization took at least five days longer to bolt than the controls at ambient temperature and the other vernalisation treatments. The expression of *DtFLC* throughout development was investigated for each treatment. Under ambient conditions, *DtFLC* expression went up until three weeks after germination and then fell at four weeks old. The expression then increased again up to bolting and fell after the transition to flowering had occurred (Figure 4.4a). For those plants which had been given the vernalization treatment, the pattern of *DtFLC* expression was the same as under the ambient conditions up until the vernalization treatment started. When the vernalization treatment began, the levels of *DtFLC* fell to a low level. Once low levels of *DtFLC* had been reached in all treatment lengths, the level remained low, even when plants were returned to ambient conditions (Figure 4.4). This suggests that all treatment lengths were effective in reducing the expression of *DtFLC* and keeping it at a low level.



**Figure 4.3 – Effect on bolting time of varying the length of 4°C cold treatment on *D. tenuifolia* four week old plants.**

Mean number of days to bolt for plants subjected to cold treatment at 4°C for two, four and eight weeks. The length of the vernalization treatment was subtracted from the number of days to bolt. Bars show mean ± SE (n=14 (0 weeks), n=15 (2 weeks and 4 weeks), n=13 (8 weeks)). Error bars denote the standard error of the mean. Student’s *t*-test was used to compare the number of days to bolt of each vernalization treatment (2, 4 and 8 weeks) against the ambient conditions (0 weeks). There was little or no significant difference in number of days to bolt between vernalization treatments and ambient conditions ( $p < 0.05$ ).

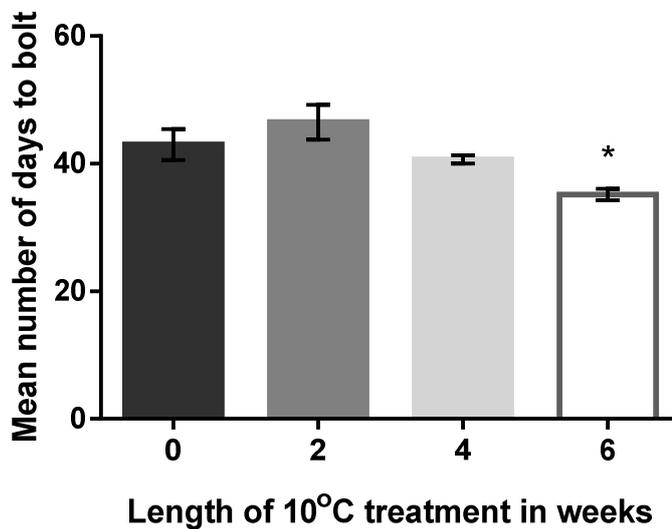


**Figure 4.4 – Expression of *DtFLC* over development in vernalization treatments of different lengths.**

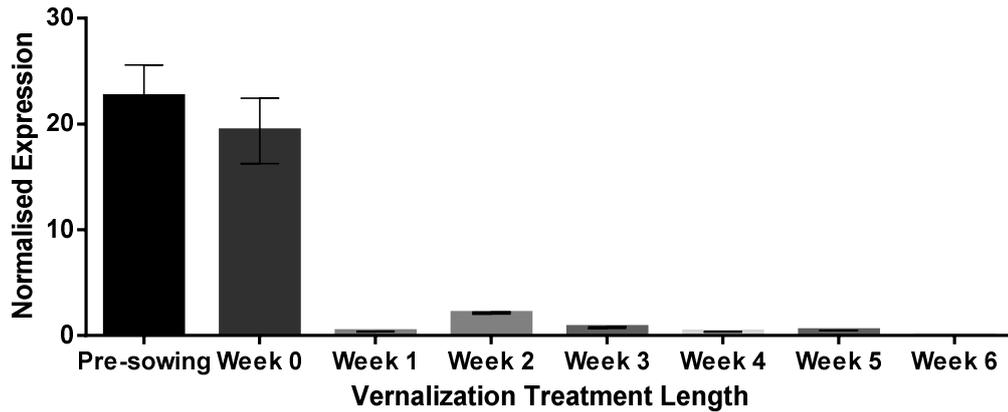
a) Change in levels of *DtFLC* expression over the course of the experiment under ambient conditions, b) two weeks vernalization treatment at 4°C, c) four weeks vernalization treatment at 4°C, d) eight weeks vernalization treatment at 4°C. Error bars denote the standard error of three technical replicates. Bolting time of plants shown by pink bar and period of 4°C vernalization shown in blue.

#### 4.3.1.3. Seed vernalization at 10°C

The effect of vernalizing seed at the higher temperature of 10°C on seed was also investigated. There was no significant difference in the time it took the plants to bolt between those kept under ambient conditions compared to those which were treated at 10°C for two and four weeks (Figure 4.5). Seeds experiencing six weeks of cold treatment bolted eight days earlier than under ambient conditions, which is statistically significant ( $p < 0.0124$ ). The expression of *DtFLC* was measured in pre-sown seed, after one week at ambient conditions and after various lengths of vernalization treatment at 10°C. Figure 4.6 shows that the vernalization treatment dramatically reduces the expression of *DtFLC* by ~10 fold within one week compared to the control ambient conditions. It then stays low up to week six.



**Figure 4.5 – Effect on bolting time of varying lengths of cold treatment at 10°C on *D. tenuifolia* seeds.** Mean number of days to bolt of plants derived from seed subjected to cold treatment at 10°C for two, four and six weeks. Bars show mean  $\pm$  SE (n=6). Error bars denote the standard error of the mean. Student's *t*-test was used to compare the number of days to bolt of each vernalization treatment (2, 4 and 6 weeks) against the ambient conditions (0 weeks). \*Statistical significance of  $p < 0.0124$ .

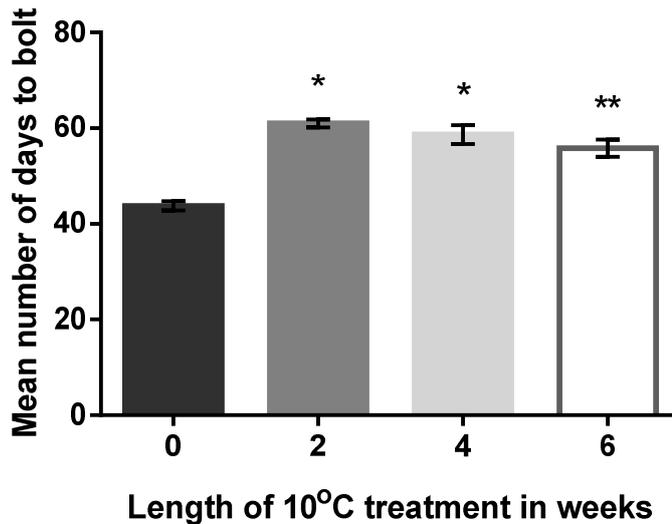


**Figure 4.6 – Effect of the different vernalization treatments at 10°C on expression of *DtFLC* in *D. tenuifolia* seed.**

Expression of *DtFLC* throughout vernalization where seeds were subjected to cold treatment of 10°C for between one and six weeks, with expression levels examined at the end of each week of treatment. Error bars denote the standard error of three technical replicates.

#### **4.3.1.4. Plant vernalization at 10°C**

Plants at four weeks old were subjected to vernalization at 10°C for two, four and six weeks. The number of days each plant underwent each treatment was recorded. This experiment showed that there is a significant difference in bolting time between the vernalized plants, and those kept at ambient temperature for the length of the experiment (Figure 4.7). Unexpectedly, the vernalization treatment significantly increased the time the plants took to bolt, 12 days longer with six week treatment ( $p < 0.0003$ ), 15 days longer with four weeks of treatment ( $p < 0.0001$ ) and 17 days longer with two weeks of cold treatment ( $p < 0.0001$ ) compared to those in ambient conditions, despite the length of the vernalization treatment being deducted from the total number of days to bolt (Figure 4.7). Unfortunately, it was not possible to investigate the expression of *DtFLC* in this experiment due to the poor quality of the RNA extracted from many time points across each treatment and across the experiment.

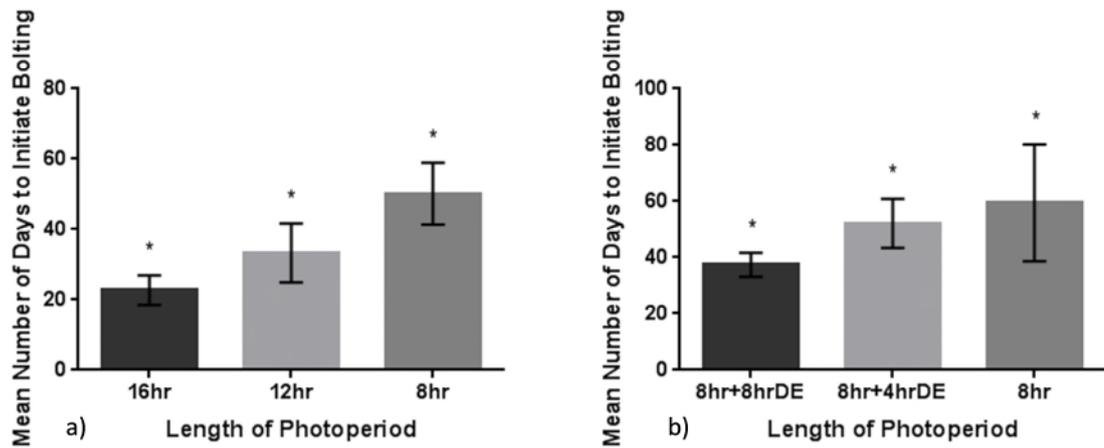


**Figure 4.7 – Effect on bolting time of varying the length of 10°C cold treatment on *D. tenuifolia* four week old plants.**

Mean number of days to bolt for plants subjected to cold treatment at 10°C for two, four and six weeks. Bars show mean  $\pm$  SE (n=5 (0 weeks), n=6 (2 weeks, 4 weeks and 6 weeks)). Error bars denote the standard error of the mean. Student’s *t*-test was used to compare the number of days to bolt of each vernalization treatment (2, 4 and 6 weeks) against the ambient conditions (0 weeks). \*Statistical significance of  $p < 0.0001$  \*\* Statistical significance of  $p < 0.0003$ .

#### 4.3.2. Photoperiodic experiments

Three different photoperiods were tested to investigate the number of days to initiate bolting in rocket under different day lengths. The experiment was repeated twice, once with fluorescent bulbs providing all the light for each of the photoperiod lengths and the second with fluorescent bulbs providing 8 hours of light and incandescent (day extension) bulbs providing the extra 4 and 8 hours light in the 12 hour and 16 hours photoperiods respectively (Table 4.1). The data show that the plants respond differently to each photoperiod, with shorter photoperiods delaying bolting (Figure 4.8). In experiment 1, the plants in each of the photoperiods initiated bolting with a significant difference in the number of days between each light period length ( $p < 0.05$ ) (Figure 4.8a). Also, all plants in the 12 hour and 16 hour photoperiods were able to bolt, but three out of nine plants in the 8 hour photoperiod had not initiated bolting by the end of the experiment (70 days).



**Figure 4.8 – Effect on bolting time of varying photoperiod lengths in *D. tenuifolia* plants.**

Mean number of days to initiate bolting for plants grown in 8 hour, 12 hour and 16 hours photoperiods. a) Photoperiod experiment 1 with light provided by fluorescent bulbs. Bars show mean  $\pm$ SD (n=9 (16hr+12hr), n=6 (8hr)). b) Photoperiod experiment 2 with light provided by fluorescent bulbs and day extension incandescent bulbs. Bars show mean  $\pm$ SD (n=5 (8hr+8hr DE, 8hr+4hr DE), n=6 (8hr)). Error bars denote the standard deviation. Ordinary one-way ANOVA showed a significant difference (\*) between all photoperiod lengths on the initiation of bolting time ( $p < 0.05$ ).

In experiment 2, the data showed a similar trend, that the longer the photoperiod, the shorter the time it takes the plants to initiate bolting, and the bolting data in each photoperiod length is significantly different from the next ( $p < 0.05$ ) (Figure 4.8b). Of the 12 plants in the 8 hour photoperiod, only one didn't initiate bolting by the end of the experiment at 100 days, but there was a range of 55 days between the initiation of bolting of the first and the last plants. This is a much greater range than observed in experiment 1, but maybe due to experiment 2 running for an extra 30 days. When comparing between experiments 1 and 2, there is an average delay in initiation in experiment 2 of 15 days under a 16 hour photoperiod, of 19 days under a 12 hour photoperiod and of 10 days under an 8 hour photoperiod. It was expected that a difference would be seen between the two experiments for the 12 and 16 hour photoperiods due to the effect of the day extension bulbs compared to the fluorescent bulbs, but here a 10 day difference has also been observed between the 8 hour photoperiods. This was unexpected as the experimental design was the same. The two experiments were run three months apart, but used the same three controlled environment cabinets with the same conditions, with the exception of the day extension bulbs. The seed line used was identical and the germination rate was comparable. The difference seen therefore may be due to a difference in timing of starting the plants in the appropriate light regime, or the number of days at which the plants were transplanted from p24 cells to 5 inch pots. Taking the discrepancy into

account, the difference in the number of days to initiate bolting in the 12 hour photoperiod is nine days, and five days in the 16 hour photoperiod. This could be due to the lower light levels produced by the incandescent bulbs providing the day extension in experiment 2 compared to the fluorescent bulbs used in experiment 1. The use of incandescent bulbs would reduce the rate of growth and photosynthesis in the plants for the time of exposure in each photoperiod.

## **4.4. Discussion**

### **4.4.1. Vernalization experiments**

In many plant species, vernalization conveys the ability of the plant to flower and can greatly reduce the time it takes to flower (Guo *et al.*, 2004; Hackett and Hartmann, 1967; Jung and Mueller, 2009; Strange *et al.*, 2011). Prior to this set of experiments it was not known whether rocket (*D. tenuifolia*) has a vernalization requirement, or if it has, at which developmental stage and temperature the response is seen. Our data shows that rocket does not respond to vernalization at 4°C, although after one week of cold treatment as seed, there is a decrease in the mean number of days to bolt. However, it is not enough to satisfy the vernalization requirement as the plants which had two and four weeks of cold treatment as seed, bolted three days and seven days later than the one week treatment, respectively (Figure 4.1). In *Arabidopsis*, Sanchez-Bermejo *et al.* (2012) explored the response of three accessions for their response to vernalization. Their results showed that in the vernalization sensitive accession, there was a big difference in flowering time between the non-vernalized plants and the early flowering vernalized plants. In rocket, although the plants which derived from seed treated with one week of cold treatment had some which bolted more quickly than the non-vernalized; the difference in the number of days was small. In the 10°C seed vernalization experiment, after six weeks of cold treatment, there was also a reduction in the number of days to bolt compared to those at constant ambient temperatures. However, similarly to the 4°C seed vernalization experiments, the difference in the number of days is only small, suggesting the effect is negligible (Figure 4.5). When the *DtFLC* expression levels in the seed are examined in both seed vernalization experiments, the results show that the levels drop after one week of cold treatment, and remain low for the further weeks of vernalization (Figure 4.2 and Figure 4.6). This suggests that the cold treatment is

effective in reducing the *DtFLC* levels but the bolting data does not reflect this. The experiments testing the vernalization at 4°C and 10°C of four week old plants showed that there is no decrease in the mean number of days to bolt in response to any of these treatments (Figure 4.3 and Figure 4.7). In some cases, the vernalized plants can take longer to bolt than those under constant ambient conditions (Figure 4.1 – Week 4, Figure 4.3 – Week 8, Figure 4.5 – Week 2 and Figure 4.7). The 4°C plant vernalization experiment was conducted in the glasshouse in summer-autumn, so those plants returning to ambient conditions after the vernalization treatments would have experienced different conditions, especially the eight week cold treated plants, as these were transferred four weeks later into the glasshouse compared to the four week vernalization treated plants. This meant that they will have been exposed to a much shorter natural day length as the season's changed and cooler temperatures, even though day extension lighting and heating was provided. This may account for the increase in number of days to bolt. The 10°C plant vernalization experiment was conducted entirely in controlled environment cabinets (Versatile plant growth MLR-352, Panasonic Co. Ltd) so the change in season would have had no effect. All plants in this experiment were delayed in bolting when treated with 10°C vernalization despite the length of treatment. This suggests that the cold treatment is having adverse effects on the plants, perhaps slowing growth which the plants take time to recover from even when returned to ambient conditions, slowing the transition to flowering. It is unfortunate that there is no *DtFLC* expression data for this experiment as this may have been informative. The *DtFLC* expression data has shown that there is a reduction in the level of *DtFLC* in response to vernalization at both 4°C and 10°C for treated seeds and at 4°C in 4 week old plants after just one week, and the low levels are maintained at all treatment lengths even when returned to ambient conditions. This suggests that rocket may have genes homologous to *VIN3*, *VRN1* and *VRN2* in Arabidopsis that reduce the expression of *FLC* (Sung and Amasino, 2004), and maintain the low levels of *FLC* after vernalization (Gendall *et al.*, 2001; Levy *et al.*, 2002). However, the reduction of *DtFLC* expression during and post-vernalization treatment is not reflected in the bolting data, suggesting that the pathway between *FLC* and flowering in rocket may not be the same as Arabidopsis. Further experiments could explore the response of *DtFLC* to vernalization by subjecting the Arabidopsis complementation lines (section 3.3.3.2), where *DtFLC* is constitutively expressed by the 35S promoter, to cold treatment.

The resulting flowering phenotype should reveal whether the gene isolated as *DtFLC* functions to confer a vernalization response in the *flc-3* mutant as observed in similar *Arabidopsis* experiments (Michaels and Amasino, 2000).

The vernalization experiments undertaken in this study are pioneer experiments and would need to be replicated with a larger number of plants with all stages of the experiments undertaken in controlled environment cabinets. The data collected suggests that rocket is not able to respond to vernalization as seeds or four week old plants. As a Mediterranean plant, a vernalization requirement in rocket may not be expected as it is a rapid cycling plant. *Arabidopsis* accessions which are rapid-cycling have loss-of-function mutations in *FRI* so have low levels of *FLC* and vernalization has little effect (Johanson *et al.*, 2000). However, the results in rocket under constant ambient temperatures have high levels of *FLC* and yet still flower quickly. This suggests it may have a functional *FRI* homologue. This will need to be tested by isolating the *FRI* gene in rocket, sequencing it and looking at expression patterns.

#### **4.4.2. Photoperiodic experiments**

The experiments carried out in this study examined the number of days to initiate bolting of plants grown in different lengths of photoperiod. Two different light conditions were used, experiment 1 with fluorescent bulbs lighting the plants for the length of the photoperiod, and experiment 2 where fluorescent bulbs were used for the first 8 hours of light and a low light day extension using incandescent bulbs to extend the day by 4 hours for the 12 hour photoperiod and 8 hours for the 16 hour photoperiod. Across both experiments, the difference between photoperiod lengths significantly affected the number of days to initiate bolting. It was observed that the longer the photoperiod, the quicker the plants initiated (Figure 4.8a and Figure 4.8b). The use of day extension bulbs in experiment 2 for the 12 and 16 hour photoperiod delayed the initiation to bolting by nearly 10 days (Figure 4.8a) compared to experiment 1 in which the fluorescent bulbs provided light for the whole photoperiod (Figure 4.8b), when the number of days difference in the 8 hour photoperiod is removed. This was to be expected because of the difference in light spectrum each bulb type provides. Fluorescent bulbs provide high blue (450-495 nm) light which is important for the stabilisation of CO through *CRY1/CRY2* enabling flowering to be promoted (Devlin, 2002; Valverde *et al.*, 2004). Incandescent bulbs provided a

much higher level of far-red light (700-800 nm) (Runkle *et al.*, 2012; Thimijan and Heins, 1983) which leads to the stabilisation of CO (Valverde *et al.*, 2004). However, incandescent lamps emitted a lower quantity of light, which may have slowed the floral transition. Switching to the incandescent bulbs to extend the long day provided some light, but not enough red and blue light to stabilize CO and induce the expression of FT, leading to a delay in flowering (Thomas, 2006). It would be useful to repeat these experiments and collect leaf material over a diurnal time course close to the initiation of bolting as done in Chapter 3. This would allow analysis of the effect the photoperiod length is having on the expression of key photoperiodic pathway genes such as *CO*, *GI* and *FT*. This would shed further light on what is happening genetically in response to the environment in rocket and how it compares to what has been observed in *Arabidopsis*.

The evidence from these experiments suggests that rocket is a facultative long-day plant as it flowers most quickly under long day conditions but is able to flower, although more slowly, under shorter day lengths as well. This knowledge could provide a way for growers who cultivate rocket under glass, to delay the onset of flowering by switching to short day conditions close to their usual harvest date when the vegetative growth is of a marketable size and quantity.

## **Chapter 5. Production and Identification of Late Bolting Lines in Rocket**

### **5.1. Introduction**

This project aimed to produce an ethyl methanesulphonate (EMS) population of rocket from which late bolting lines could be selected. *D. tenuifolia* seed resulting from four generations of single seed descent (SSD4) was treated with two different concentrations of EMS. Seed was sown and surviving plants were selfed and M<sub>2</sub> seed collected. Late bolting plants were then selected from the M<sub>2</sub> generation and backcrossed to wild type. The resulting seed was also selfed and late bolting plants selected in the subsequent generation. At generation M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub>, two lines were selected and whole genome sequencing performed on bulk segregants in order to identify SNPs co-segregating with the late bolting phenotype.

#### **5.1.1. Mutagenesis by ethyl methanesulphonate**

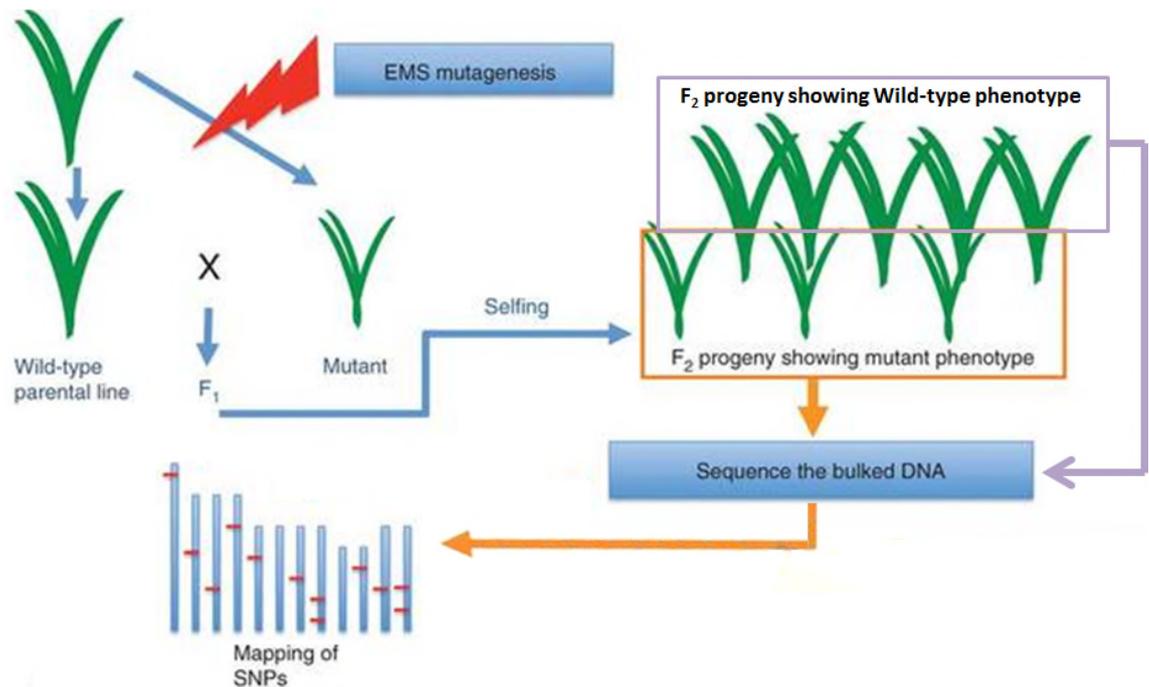
EMS is a chemical mutagen which causes changes in the DNA of the organism which is subjected to it. It is a more effective method of inducing mutations than ionizing radiation due to its alkylating ability (Brock, 1976). EMS functions by donating an alkyl residue to guanine or thymine nucleotides causing a pairing error, allowing a change in the sequence from G/C to T/A or A/T to G/C (Kim *et al.*, 2006b; Lai *et al.*, 2004). The aim of the treatment is to allow many of these changes to occur randomly across the genome giving a high mutation load and therefore an increased likelihood of induced changes leading to beneficial phenotypes. However, EMS is lethal in high concentrations so a compromise must be made to offset the high mutation load against lethality. It also affects fertility in the subsequent generation, so this too must be considered (Stephenson *et al.*, 2010). Both the concentration of dose and treatment length have an effect, as does variation in the susceptibility of seed of different plant species, and even variability between EMS batches, making it difficult to standardise the procedure without a lot of testing (Kim *et al.*, 2006b; Lai *et al.*, 2004).

#### **5.1.2. Utilising EMS to identify desired phenotypes**

EMS-based mutagenesis is now a widely used technique for random mutagenesis with applications found in many plant species such as *A. thaliana* (Brock, 1976; Kim

*et al.*, 2006b), *Oryza sativa* (Abe *et al.*, 2012), *B. rapa* (Stephenson *et al.*, 2010), *Capsicum annum* L. (Arisha *et al.*, 2015), *Solanum lycopersicum* (Gady *et al.*, 2009) and *D. tenuifolia* (Kenigsbuch *et al.*, 2009). This is because it is not classed as a transgenic approach so any varieties bred from EMS treated lines are not subject to GM regulations (Till *et al.*, 2006). Kenigsbuch *et al.* (2009; 2014) successfully used EMS to induce mutations and select for late bolting and slow senescing rocket plants. Their late bolting candidate ‘Rock-Ad’ was two weeks later than the wild-type, showing that EMS treatment is a viable approach to obtain significant delay of bolting in rocket. In our work we also aim to analyse and sequence the late bolting lines to identify the molecular basis of the delayed bolting, something that was not investigated in Kenigsbuch *et al.* (2009; 2014).

Targeted induced local lesions in genomes (TILLING) populations in numerous species have been produced, where large numbers of seed are subjected to EMS treatment and screened for various phenotypes, whilst structured pooling of DNA from individuals in the population and screening these pools with gene-specific primers and high throughput sequencing can enable plants to be identified carrying SNPs in specific genes (Stephenson *et al.*, 2010; Till *et al.*, 2006; Tsai *et al.*, 2013). This can provide a fast way of identifying where SNPs are present and in which plant line this occurs. This technique, however, relies on the ability to create a large population with which to screen, which will not be possible in this project. A second approach to fast SNP discovery was described by Abe *et al.* (2012) in rice, called MutMap. The EMS treated plants were screened for the desired phenotype. These were then backcrossed once to the wild type control and selfed, producing a segregating F<sub>2</sub> population. Those plants which show the desired phenotype are pooled together and sequenced along with the wild type control and SNPs discovered by comparing the two genomes (Figure 5.1). A similar approach to this will be used in this project as it is expected that only one backcross and selfing generation will be possible in the project timeline. The method will be adapted by pooling together the F<sub>2</sub> progeny which show the desired late bolting phenotype and F<sub>2</sub> progeny which show the wild type bolting phenotype and sequence these rather than the wild type control. The SNPs identified in the sequencing should therefore reveal those only which are linked to the late bolting phenotype as other non-linked SNPs should be present evenly in both pools (Figure 5.1).



**Figure 5.1 – MutMap method for fast SNP discovery in an EMS population.**

Plants are mutagenized with EMS, and desired mutant phenotype plants selected and backcrossed to the wild-type parental line producing the F<sub>1</sub> generation. These are selfed, producing the segregating F<sub>2</sub> generation. F<sub>2</sub> progeny showing the mutant phenotype (orange box) are pooled and sequenced alongside the pooled wild-type phenotype F<sub>2</sub> progeny (purple box) and SNPs are mapped and linked to the mutant phenotype. Figure adapted from Abe *et al.* (2012).

## 5.2. Methods

### 5.2.1. EMS treatment

EMS was acquired from Sigma Aldrich and used to treat *D. tenuifolia* seeds. A trial experiment was set up to test different concentrations of EMS on genebank stock seed (section 2.1.2.2). Once the concentration and method had been optimised, two batches of 2000 *D. tenuifolia* SSD4 were treated with EMS (section 2.1.6.3), scored for germination and planted into the field (section 2.1.7.2). Self-seed was collected from individuals and sown out as the M<sub>2</sub> generation into Spanish polytunnels. Late bolting plants were then selected for backcrossing and selfing (section 2.1.6.4). Seed for M<sub>3</sub> generation was also sown into the field to assess segregation of the late bolting phenotype (section 2.1.6.6), and M<sub>2</sub>\_BC<sub>1</sub> seed was sown in the glasshouse to produce self-seed from late bolting plants. Fourteen late bolting lines were selected for sowing of M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub> to assess segregation of late bolting to wild type bolting plants (section 2.1.6.5). Of these, four were chosen and re-sown with 30 seeds per line ready for DNA extraction and whole genome sequencing.

### **5.2.2. Whole genome sequencing**

DNA was extracted from late bolting and wild type bolting plants from lines M2\_723\_4\_BC1\_f1\_2 and M2\_723\_4\_BC1\_f1\_3 using the CTAB method (Stewart and Via, 1993) and finally resuspended in DEPC-treated water. DNA was quantified using Qubit Fluorometer (Invitrogen) and pooled into two groups, one for the late bolting plants and one for the wild type bolting plants. A library was made by the Genomics Facility (University of Warwick) for each pool using a Nextera® DNA sample preparation kit (Illumina). The libraries were submitted to the Wellcome Trust for HiSeq whole genome sequencing.

### **5.2.3. Pipeline for genome assembly and SNP identification**

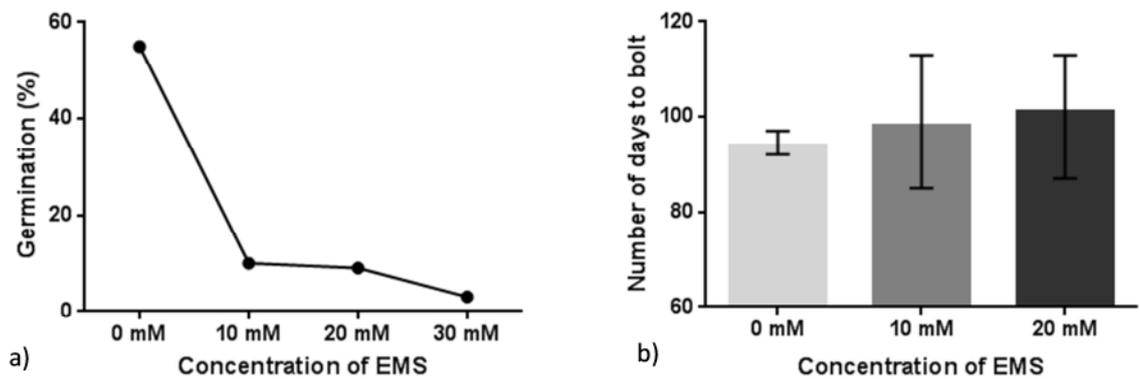
Once the HiSeq whole genome sequencing had been performed, the data was returned for analysis. The quality of the raw reads was examined using FASTQC followed by trimming the ends of the reads using Trimmomatic to remove low quality end sequence and adapter contamination. The trimmed reads were paired into an equal number of forward and reverse pairings and FASTQC was performed again to ensure a high quality score. The depth of sequencing was calculated using the formula  $N * L/G$  where N=number of reads, L=read length and G=genome size. The genome size for *D. tenuifolia* was estimated using the known genome size of *D. erucoides* at 632 Mbp (Arumuganathan and Earle, 1991). The DiscoSNP programme was used to compare the SNPs present between the late bolting pool sample and the wild type bolting pool sample filtering for those with a depth of coverage of at least 30, and where the SNPs were of the expected mutation type due to the EMS treatment (A/G, G/A, C/T, T/C). The SNPs were then further filtered to find those which were present only in one sample but not the other (1,-1 and -1,1). Single contigs (unitigs) containing the SNPs were BLASTed to identify which genes they were homologous to. A *de novo* assembly was also built using RAY with a K-mer value of 31. This enabled the search of the SNP sequences against the whole genome to find which contigs they lie in and therefore a more accurate BLAST search could be done, as more sequence data is available.

## 5.3. Results

### 5.3.1. EMS mutagenesis

#### 5.3.1.1. EMS trial data

Germination data from EMS trial 1 showed that the higher the concentration of EMS, the lower the number of seed that germinated (Figure 5.2a). This result was to be expected and reflects what is reported in the literature (Stephenson *et al.*, 2010; Thurling and Depittayanan, 1992).

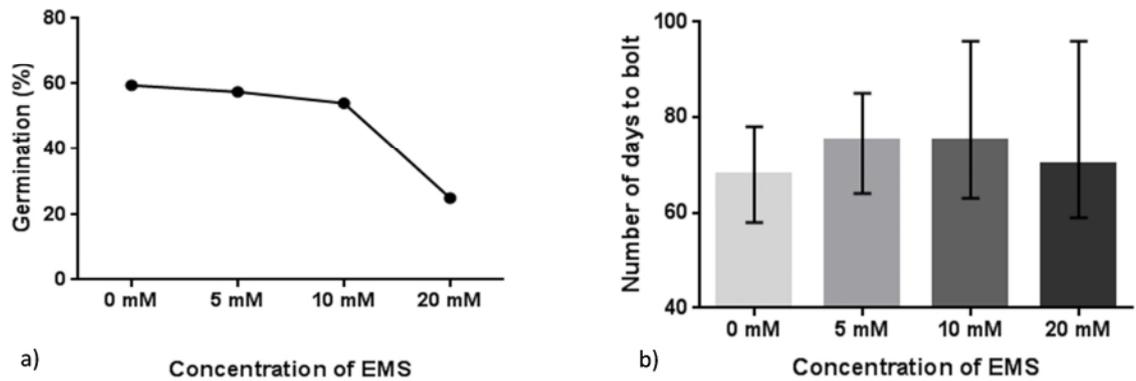


**Figure 5.2 – Effect of EMS concentration on seed germination and number of days to bolt in Trial 1.**

Rocket seeds were treated at three concentrations of EMS along with the negative control (0 mM). a) Percentage of seed which germinated was recorded after 7 days for each EMS concentration. b) Mean number of days to bolt at each EMS concentration. Error bars show the range of number of days to bolt of individuals at each EMS concentration.

Whilst a few seeds treated with 30 mM EMS germinated, they did not survive long enough to transplant due to post-germination lethality of the EMS treatment. Bolting data was obtained for the seedlings that survived the EMS treatments at 10 mM and 20 mM. The data shows that there was an increase in the range of bolting times across the population at each concentration compared to the non-treated sample (0 mM) (Figure 5.2b). A second trial was therefore performed using lower concentrations of EMS such as 5 mM, 10 mM and 20 mM.

The germination results of the second trial showed that using a lower concentration of EMS meant that a higher percentage of seed germinated (Figure 5.3a). However, the percentage germination at 10 mM and 20 mM is higher than that in the first treatment which may reflect the difference in the genotype of the seed batch used, or the inconsistency of the EMS batches used (Stephenson *et al.*, 2010).



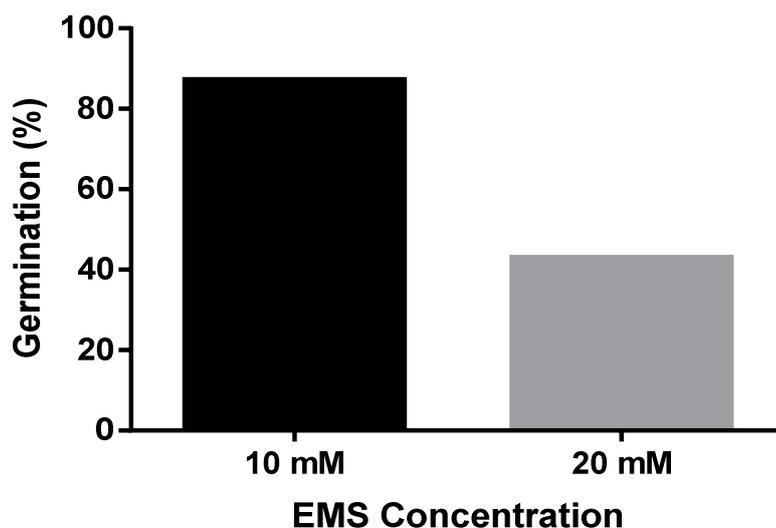
**Figure 5.3 – Effect of EMS concentration on seed germination and number of days to bolt in Trial 2.**

Rocket seeds were treated at three concentrations of EMS along with the negative control (0 mM). a) Percentage of seed which germinated was recorded after 7 days for each EMS concentration. b) Mean number of days to bolt at each EMS concentration. Error bars show the range of number of days to bolt of individuals at each EMS concentration.

Bolting data was recorded for each plant in trial 2. The results reflect what was observed in trial 1, which is that as the concentration of EMS increased, the range of the number of days to bolt at each treatment increased compared to the untreated control population (Figure 5.3b). This suggests that the EMS has introduced a variety of mutations which affect the bolting time, and that the greater the mutagenized load, the greater the spread of bolting times in the population.

### **5.3.1.2. Production of $M_1$ plants and $M_2$ seed**

The large scale EMS treatment was performed on 4000 SSD4 *D. tenuifolia* seeds (section 2.1.2.1). With the evidence collected from the trial experiments, two concentrations of EMS were used (10 mM in experiment 1 and 20 mM in experiment 2) to balance the EMS lethality rate with a high mutation load. Germination was recorded for each experiment (Figure 5.4). The germination at each concentration is higher than in either of the previous trials, highlighting how variable the mutagenesis effects can be even when precautions are taken to repeat exactly what was done before. Much larger numbers of seeds were used in this experiment, however, so it may be more representative of the EMS effects than the previous smaller trials.



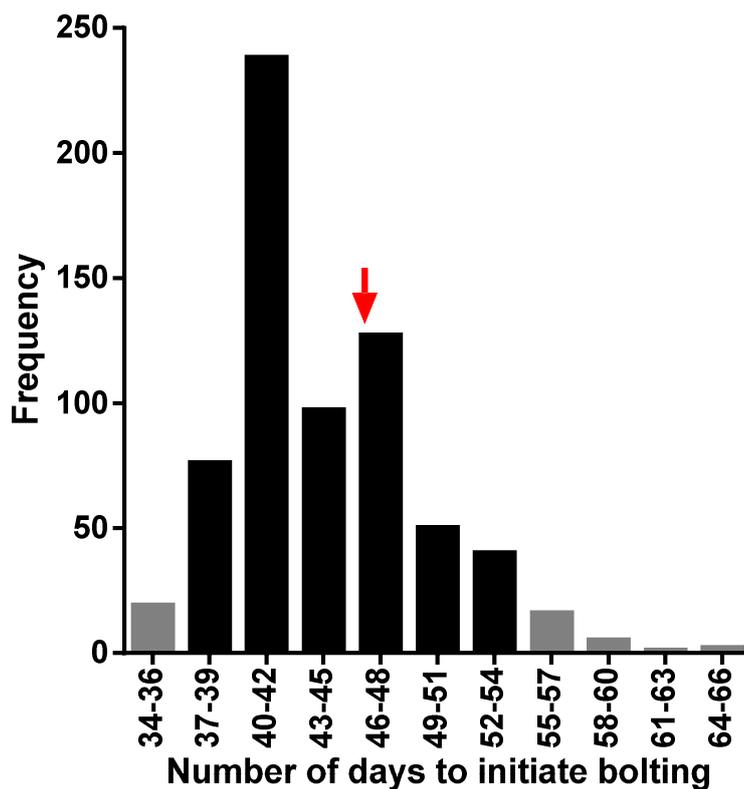
**Figure 5.4 – Effect of EMS concentration on seed germination in EMS experiments 1 and 2.**

2000 rocket seeds were treated at two concentrations of EMS, 10 mM in experiment 1 and 20 mM in experiment 2. Germination (%) was recorded after seven days

EMS is known to adversely affect fertility, and as the season was unfavourable for growth of the  $M_1$  rocket plants, flies were added to the population in perforated plastic bags to aid pollination and promote  $M_2$  seed set (section 2.1.4.1). At the season end, the plants were cut and dried before threshing. 1127 plants in total were threshed, 631 from experiment 1 and 496 from experiment 2.

### **5.3.1.3. Screening the $M_2$ population**

Of the 1127 plants threshed, only 332 plants produced seed, 256 from experiment 1 and 76 from experiment 2. This result was not surprising as EMS treatment affects fertility and *D. tenuifolia* is a natural outcrosser so does not readily produce selfed seed (Eschmann-Grupe *et al.*, 2004). Where possible, six seeds were sown from each of the 332 lines but only 223 lines germinated and produced seedlings. EMS is known to affect germination in the subsequent generation (Kim *et al.*, 2006b), so although disappointing, this result was also to be expected. Number of days to bolt was recorded for this generation (Figure 5.5) as well as other phenotypic differences, such as plant size, leaf shape and leaf colour.



**Figure 5.5 – Bolting frequency of  $M_2$  plants.**

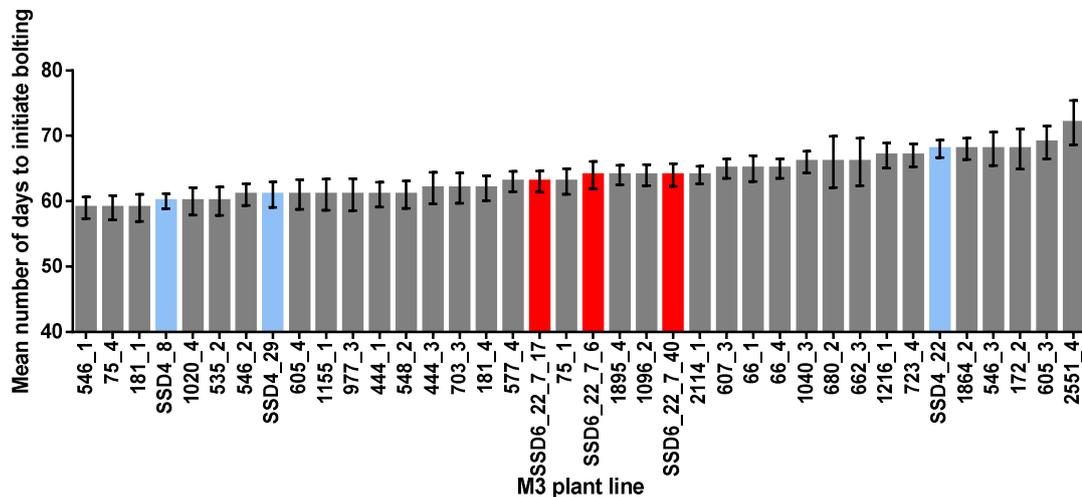
Up to four plants per  $M_2$  line were planted in Spanish polytunnels along with 94 wild type (WT) control plants from four different WT parent lines (progeny of SSD4 WT plants sown in  $M_1$  trial). Number of days to bolt was recorded when initiation of bolting occurred. Frequency of  $M_2$  plants for number of days to initiate bolting is plotted. Red arrow shows the mean number of days to initiate of the WT population (46 days). Grey bars show the number of  $M_2$  plants which initiate bolting with greater than or less than 2 standard deviations from the WT mean.

The  $M_2$  plants showed a broad range of bolting times, with 21 individuals which were later than the wild type mean (46 days) by two standard deviations. A total of 30 plants initiated bolting with at least seven days later than the WT mean which is enough of a delay to be commercially interesting. These 30 plants were dug up and moved into the glasshouse to allow controlled selfing and backcrossing to SSD6 wild type lines. However, 25 out of the 30 late bolting plants were smaller than their counterparts so they may have retarded development rather than a delay in bolting, or the small size is a pleiotropic effect segregating with the late bolting phenotype.

#### **5.3.1.4. Screening of the $M_3$ population**

Self-seed was collected from the late individual plants from  $M_2$  and eight seeds per plant were sown out along with wild type control plants of SSD4 and SSD6.

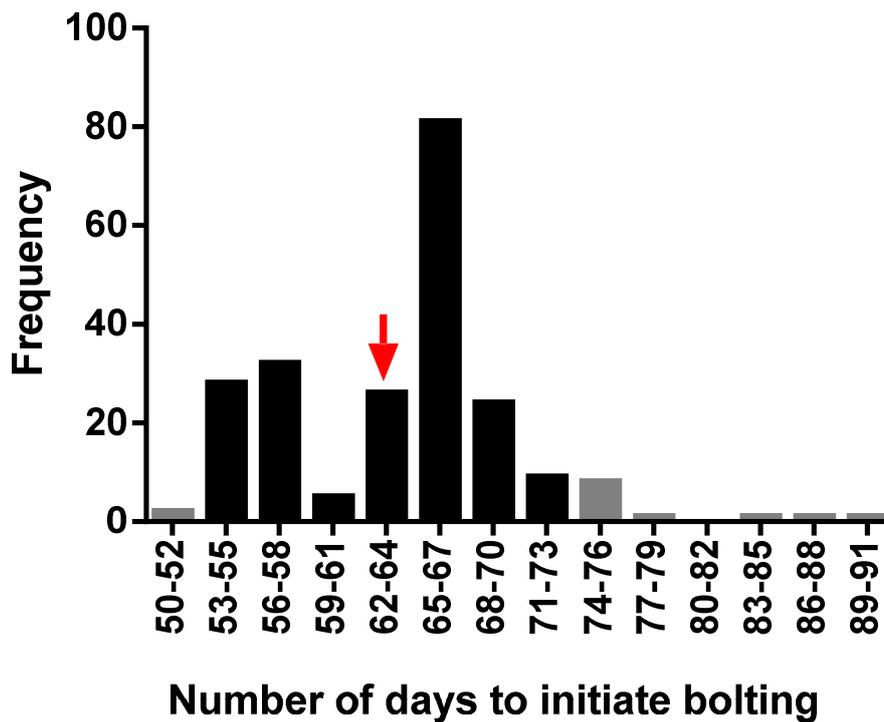
Seedlings were transplanted into the field at four weeks from sowing and screened for bolting time. This was to test whether the lines were still segregating and if the late bolting phenotype observed in the M<sub>2</sub> generation was still present in the M<sub>3</sub> generation.



**Figure 5.6 – Mean number of days to initiate bolting for EMS M<sub>3</sub> generation.**

Up to eight plants per M<sub>3</sub> line and up to ten plants per wild type control line were scored for number of days to initiate bolting. The mean number of days to initiate per line was calculated and plotted. Red bars show mean number of days to bolt for the three SSD6 wild type lines and the blue bars show the mean number of days to initiate bolting for the three SSD4 wild type lines. Error bars show the range of number of days to initiate bolting per line.

Figure 5.6 shows the mean number of days to bolt for each line in the M<sub>3</sub> generation. The difference in the number of days between the mean of the earliest and mean of the latest bolting lines is 13 days, with all six wild type lines lying in between. The error bars show the range of the number of days to initiate for individuals in each line. Some lines such as M<sub>3</sub>\_680\_2 and M<sub>3</sub>\_662\_3 have a large range of days to initiate bolting suggesting they are still segregating. The wild type lines in blue are SSD4 generation which the EMS treatment was performed on and show some variation in bolting time within and between lines. The SSD6 lines in red are still variable within each line, but there is less variation between lines than in the SSD4 suggesting these lines are more homozygous with less variability which is to be expected. The M<sub>3</sub> lines on the right-hand side of the red wild type bars are the most interesting as they show in general that there are some plants in the line which are bolting later than wild type.



**Figure 5.7 – Bolting frequency of M<sub>3</sub> plants.**

Up to eight plants per M<sub>3</sub> line were planted in the field along with 49 wild type (WT) control plants from 6 different WT parent lines (3 SSD4 and 3 SSD6). Number of days to bolt was recorded when initiation of bolting occurred. Frequency of M<sub>3</sub> plants for number of days to initiate bolting is plotted. Red arrow shows the mean number of days to initiate of the WT population (63 days = 62 days in SSD4 and 63 days in SSD6). Grey bars show the number of M<sub>3</sub> plants which initiate bolting with greater than or less than 2 standard deviations from the WT mean.

Figure 5.7 shows the number of days to initiate bolting of all the M<sub>3</sub> individual plants in the trial. The wild type mean of 63 days is highlighted with a red arrow. The grey bars show all the individuals which have bolted with greater than or less than two standard deviations from the wild type mean. Twelve plants were found to be later than the wild type by two standard deviations but six of these were recorded as small or with abnormal growth suggesting there may be other reasons for the late bolting rather than a delay in floral transition. All twelve individuals apart from the one in line M<sub>3</sub>\_703\_3 are present in the lines whose mean is greater than that of the SSD6 (red bars) wild type lines seen in Figure 5.6. The data from this generation, although complicated, helped to inform which M<sub>2</sub>\_BC<sub>1</sub> lines should be chosen as late bolting candidates.

#### **5.3.1.5. M<sub>2</sub>BC<sub>1</sub> population**

Backcrossing using bud pollination was carried out on 42 plants from the M<sub>2</sub> population using wild type SSD6 plants as the pollen donor. This included the 30 individuals identified as bolting seven days later than the wild type in the M<sub>2</sub> generation and 12 which were still later than wild type but by a smaller margin. Thirty five plants successfully set seed from the backcross events and were harvested. Three seeds per backcrossed flower were sown to produce self-seed and grown in the glasshouse. Germination of some lines was poor, but many plants also died as young seedlings so many lines were re-sown one month later. Unfortunately, all the wild type plants died so were re-sown, giving a comparative bolting phenotype for the second sowing population only. The bolting phenotype for all plants was recorded.

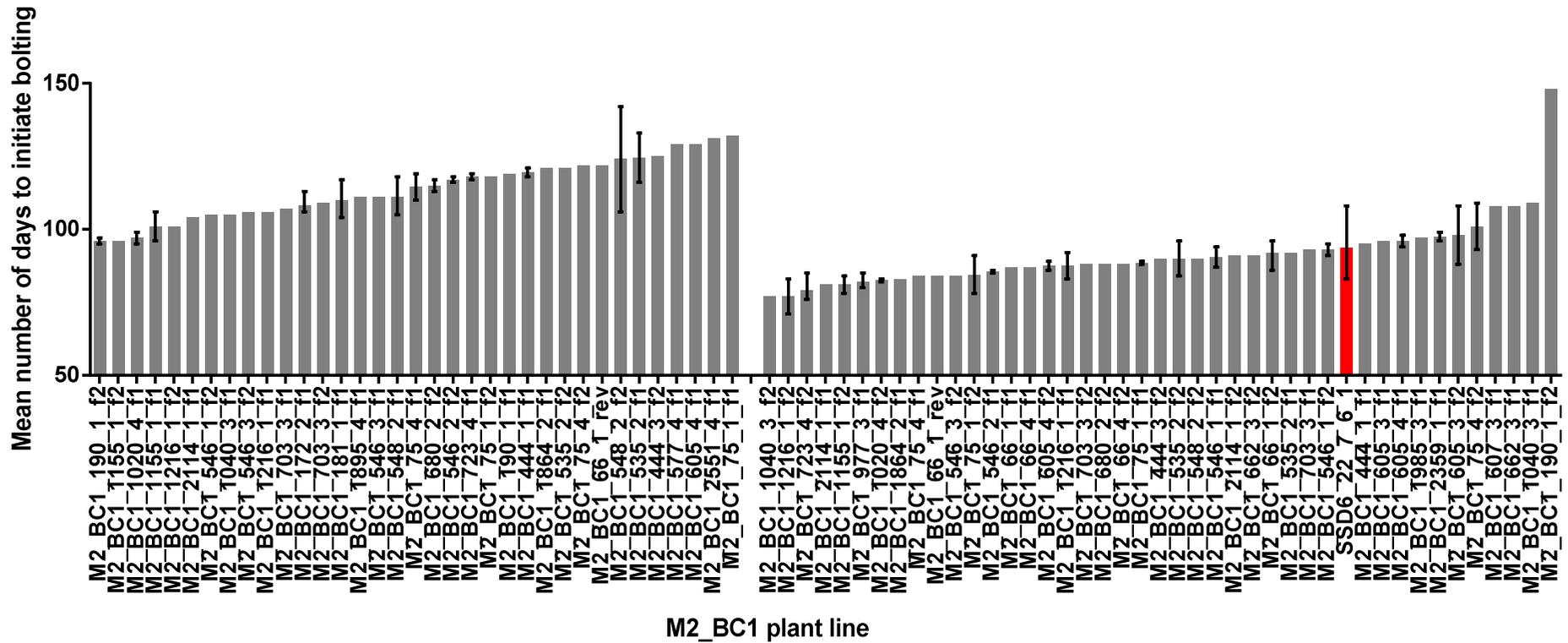


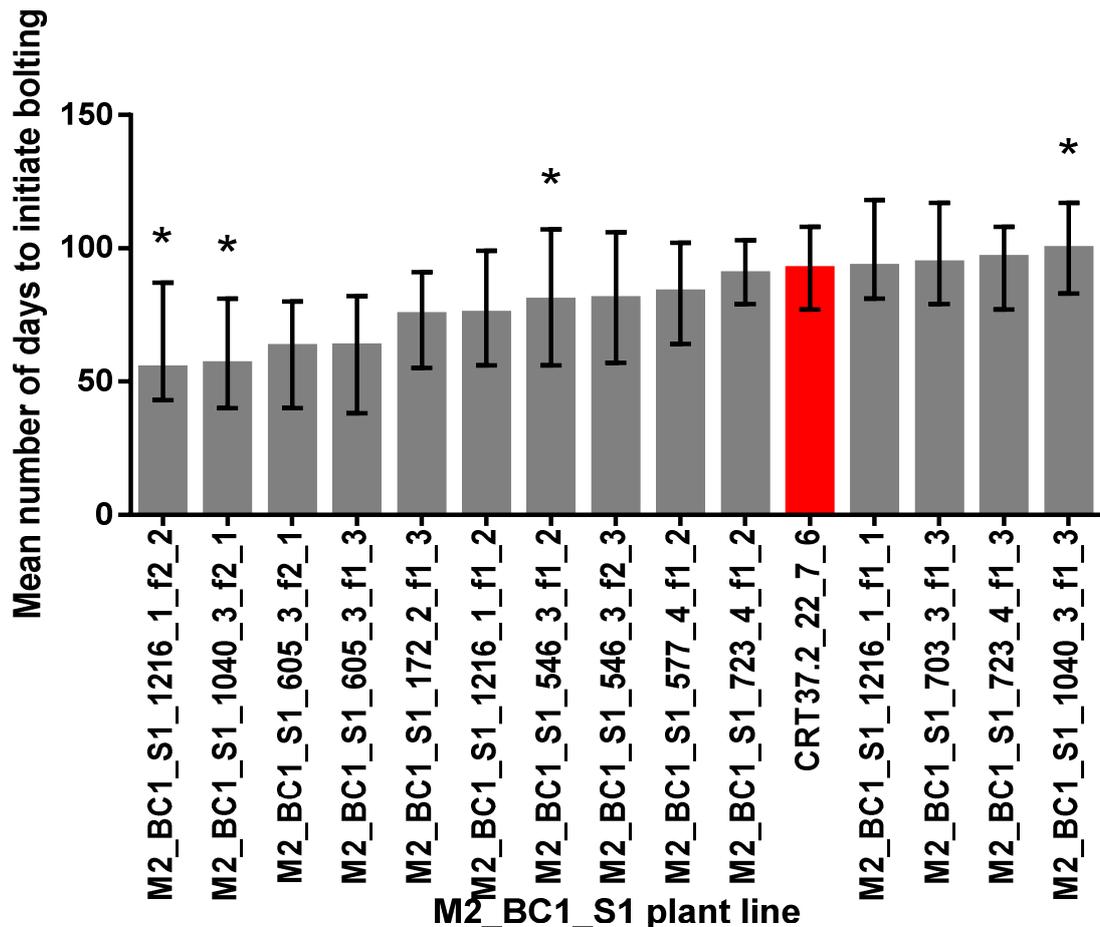
Figure 5.8 – Bolting data for M<sub>2</sub>\_BC<sub>1</sub> generation.

Mean number of days to initiate bolting for M<sub>2</sub>\_BC<sub>1</sub> lines sown a month apart. Left-hand graph shows lines sown on 03/12/2013; right-hand graph shows lines re-sown on 06/01/2014. Red bar shows wild type mean of 15 SSD6 plants. Error bars show range of bolting times in each line. Where error bars are absent, only 1 plant represents the line.

The bolting data showed that there is a range of bolting times throughout the population, with one plant (M<sub>2</sub>BC<sub>1</sub>\_190\_1\_f2) being much later than the latest wild type and retaining the later bolting phenotype (Figure 5.8). However, due to the small number of seed per cross sown and the lack of equivalent wild type plants in one of the two populations it was very difficult to draw any firm conclusions about which lines to take to the subsequent generation. Selfing by bud pollination was attempted on all plants.

#### **5.3.1.6. Late bolting segregation in M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub> population**

Twenty three M<sub>2</sub>BC<sub>1</sub> lines from nine M<sub>2</sub> lines were continued to the next generation. All of these lines were represented in the M<sub>3</sub> generation, six with a late bolting mean and two with a wild type bolting mean. Ten seeds from each of the 23 M<sub>2</sub>BC<sub>1</sub> lines were sown and nine lines did not germinate. Wild type SSD6 seed were sown out alongside the EMS lines and the number of days to initiate bolting was recorded.

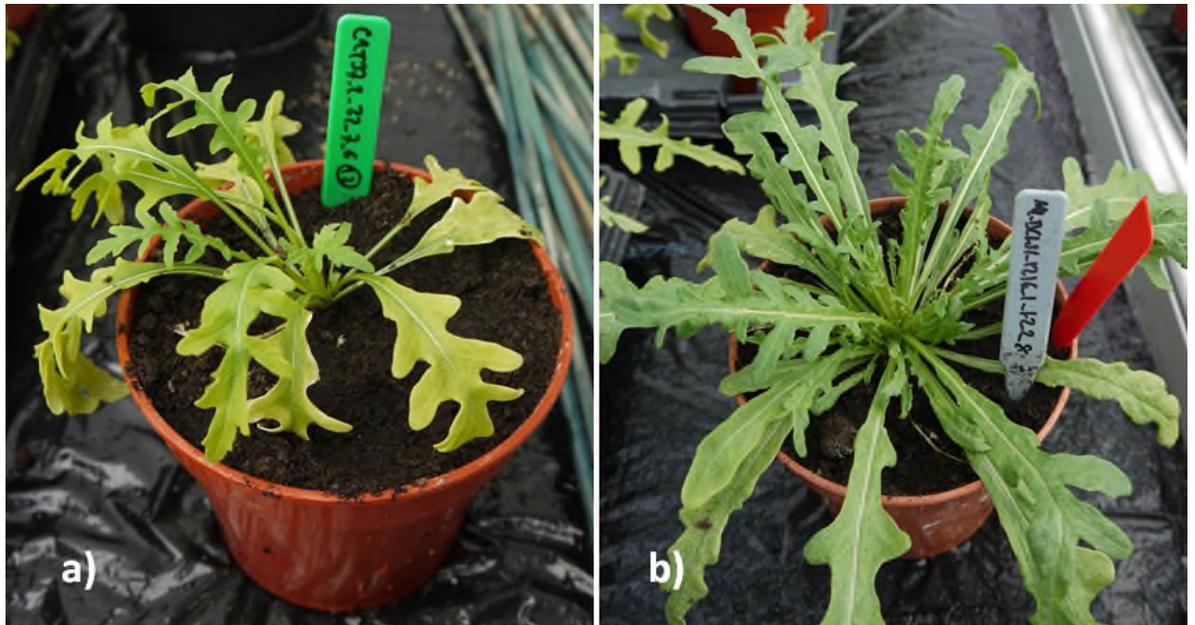


**Figure 5.9 – Bolting data for M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub> generation.**

Mean number of days to initiate bolting of each M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub> line (10 plants per line). Red bar shows wild type mean of 30 SSD6 plants. Error bars show range of bolting times in each line. Asterisks show lines where leaf abnormalities were observed.

The bolting data showed that there were three lines which had a greater mean number of days to initiate bolting than wild type, and three lines which had individuals that were later than the latest wild type plant (Figure 5.9). All lines had a range of bolting times indicating the segregation occurring within the line. However, there is still more of a spread in the bolting data for the wild type than desired or expected. Four lines contained a number of plants which are possible pleiotropic mutants as they showed a mottled, deformed leaf phenotype. These were therefore discounted from the study (Figure 5.9 and Figure 5.10). Four lines (M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>703\_1\_f1\_3, M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_2, M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_3 and M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>1216\_1\_f1\_1) were chosen for a larger segregation screen for whole genome sequencing as they had the same or later mean number of days to initiate

bolting compared to the wild type, a wide range of bolting times indicating segregation and a normal phenotype suggesting the late bolting observed is due to a delay in the floral transition.

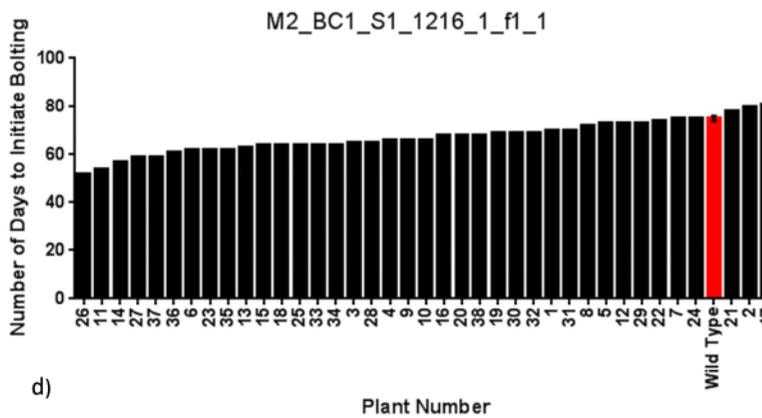
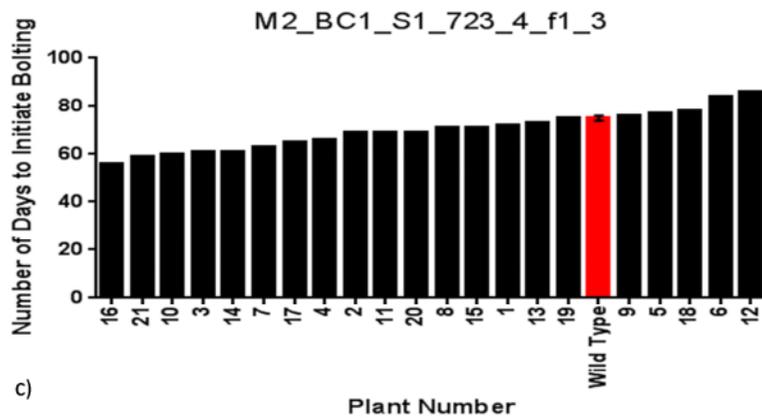
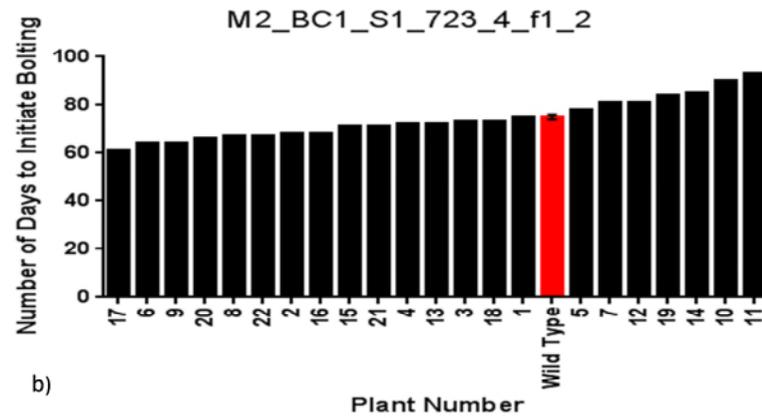
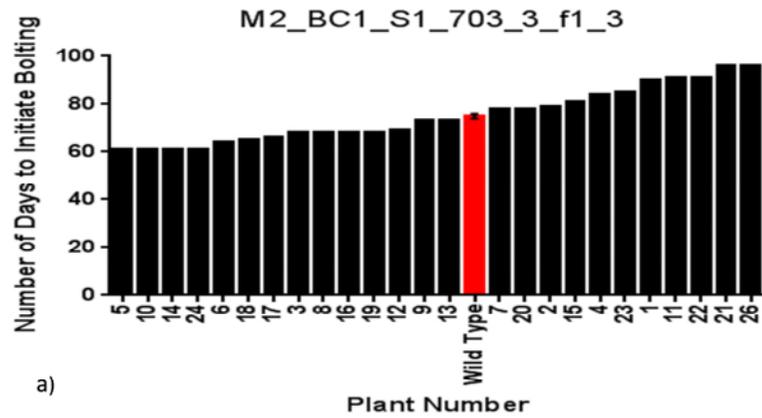


**Figure 5.10 – Normal and mottled, deformed leaf type.**

a) Wild type SSD6 plant showing a normal leaf type. b) A pleiotropic mutant M2\_BC1\_S1 plant showing a mottled, deformed leaf type. Photos taken at 62 days after sowing.

### **5.3.2. Identification of lines suitable for whole genome sequencing**

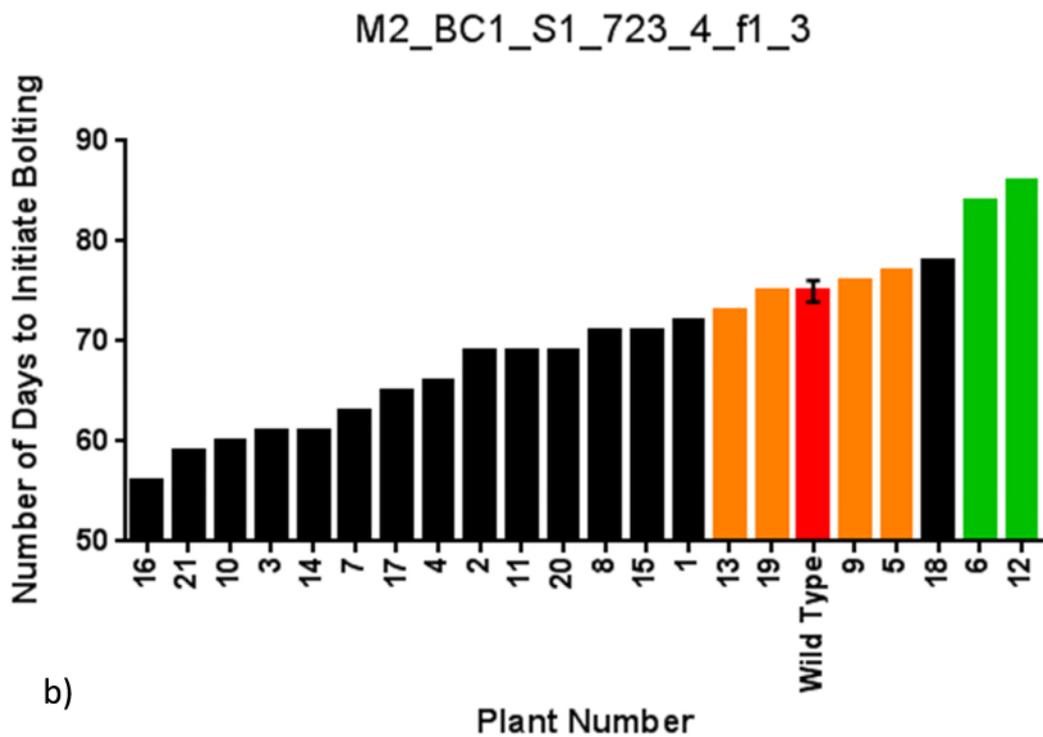
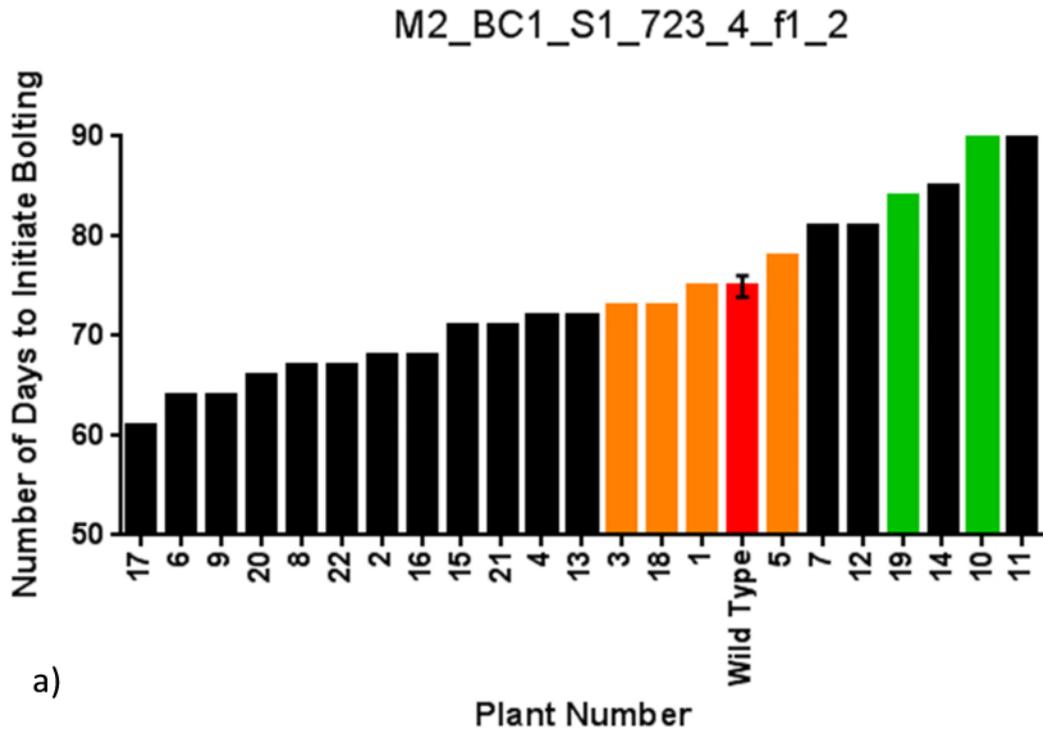
Thirty seeds of the four M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub> lines identified in the previous bolting phenotype screen were sown out into the glasshouse. Fifty SSD6 wild type seeds were also sown to act as a comparison for the bolting phenotype.



**Figure 5.11 - Bolting data for  $M_2\_BC_1\_S_1$  lines.**

Number of days to initiate bolting was recorded for each of the four lines chosen for the larger segregation screen. a) Line  $M_2\_BC_1\_S_1\_703\_3\_f1\_3$ . b) Line  $M_2\_BC_1\_S_1\_723\_4\_f1\_2$ . c) Line  $M_2\_BC_1\_S_1\_723\_4\_f1\_3$ . d) Line  $M_2\_BC_1\_S_1\_1216\_1\_f1\_1$ . Red bar shows the mean number of days to initiate bolting for the SSD6 wild type. Error bars denote the standard error. An Ordinary one-way ANOVA on these lines and wild type population showed a significant difference in mean bolting time of line  $M_2\_BC_1\_S_1\_1216\_1\_f1\_1$ , where it is shown to be earlier bolting than the other lines and wild types in the population ( $p < 0.0001$ ).

Plants in line M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>703\_3\_f1\_3 showed a good range of bolting times and of segregation, with ten individuals bolting later than the wild type mean. However, six out of these ten plants were smaller in size than normal growing plants which could account for its late bolting phenotype so this line was not considered for whole genome sequencing (Figure 5.11a). Plants in line M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_2 had seven individuals bolting later than the wild type with two having the small size phenotype. The latest bolting, non-small plant bolted 18 days after the wild type average so this line was considered for whole genome sequencing (Figure 5.11b). Line M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_3 showed a range of bolting time of 30 days with five plants bolting later than the wild type average (Figure 5.11c). No plants were found to be smaller in size than the average, but four were observed to have a rolled leaf phenotype. Despite this, the line was still considered a good candidate for whole genome sequencing. The line M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>1216\_1\_f1\_1 population had a spread of 16 days in the number of days to bolt from the quickest to the slowest. Only three plants bolted later than the wild type average and none bolted later than the latest wild type individual (Figure 5.11d). When an ordinary ANOVA was performed on all the populations in this screen, line M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>1216\_1\_f1\_1 was found to take a significantly smaller number of days to initiate bolting than the rest ( $p < 0.0001$ ). It was therefore not considered for whole genome sequencing.



**Figure 5.12 – Segregating populations of M2\_BC1\_S1 lines chosen for whole genome sequencing.** Mean number of days to initiate bolting for lines a) M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub>\_723\_4\_f1\_2 and b) M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub>\_723\_4\_f1\_3. Red bar shows mean number of days to initiate bolting for wild type SSD6 population. Green bars show late bolting plants to be pooled for whole genome sequencing. Orange bars show wild type-like bolting plants to be pooled for whole genome sequencing. Error bars denote standard error of the mean.

The two lines M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_2 and M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_3 considered for whole genome sequencing were assessed for suitable plants for late bolting and wild type-like bolting candidates. Figure 5.12 shows the late bolting plants in green and the wild type bolting plants in orange. In line M<sub>2</sub>BC<sub>1</sub>S<sub>1</sub>723\_4\_f1\_2, plants 11 and 14 were not selected due to a small size phenotype which could be contributing to the late bolting phenotype observed. Both lines originate from the same M<sub>2</sub> plant and flower which was backcrossed to wild type so share lineage and the mutation causing the late bolting will be the same in these plants. Therefore it was possible to pool the plants from both lines together for the sequencing.

### 5.3.3. SNP identification

The DNA extracted from late bolting plants (green) and wild type-like bolting plants (orange) (Figure 5.12) was pooled together and two libraries, one late and one wild type bolting, made for whole genome sequencing. The sequencing was performed and the data was analysed following the pipeline outlined in section 5.2.3. Based on the genome size estimation of 632 Mbp, the coverage of the wild type bolting sample is 46x, and 53x for the late bolting sample. However, when the *de novo* assembly was done, the genome size was estimated at ~318 Mbp. This genome size is a haploid value and 50% smaller than the estimate. The N50 value of the sequence was found to be 1.9 kb with a Kmer size of 31. 609 SNPs which differed between the wild type bolting and late bolting samples with a confidence ranking of 0.7 or more were discovered. Five of these had a confidence ranking of 1 so were investigated for their homology to any known genes (Appendix A). Three of these were found to be homologous to known genes when BLASTed. BLASTn results of the SNP 1 sequence identified three proteins with an equal amount of homology (17% coverage and 74% identity). These were *B. napus* protein C2-DOMAIN ABA-RELATED 2-like (involved in ABA sensing (Rodriguez *et al.*, 2014)), *B. rapa* SRK-54, SP11-54 genes for S receptor kinase-54 (involved in self-incompatibility (Fujimoto *et al.*, 2006)) and *B. napus* SWI/SNF complex subunit SWI3A-like (a transcription regulatory protein (<http://hintdb.hgc.jp/http/proteins/Q8W475.html>)). However, when the same sequence was run using BLASTx, no homologous proteins were found. SNP 2 was BLASTed using BLASTn where a coverage of 37% and 73% identity was found to *B. napus* 5-enolpyruvylshikimate-3-phosphate synthase gene (a key enzyme in glyphosate herbicide resistance (Klee *et al.*, 1987)), and a

14% coverage, 74% identity to *B. oleracea* var. *oleracea* gibberellin 20 oxidase 1-like gene (involved in GA flowering pathway (Hedden and Phillips, 2000)). However, no homologous genes were found using BLASTx with this sequence, or with the full contig containing this sequence. SNP 3 was found to be homologous to *B. napus* disease resistance protein TAO1-like (a target of avirulence gene B (Eitas *et al.*, 2008)) with 50% coverage and 93% identity using BLASTn. The same protein was identified when BLASTx was used for the SNP sequence and the contig it lies within. A further 604 SNPs were also investigated and each BLASTed to identify whether any are homologous to known genes. In the set of SNPs identified with a confidence ranking of 0.8, one SNP (SNP 6, Appendix C.1.6) was found to be homologous to *B. rapa* protein CENTRORADIALIS-like at a coverage of 35% and 83% identity and *A. thaliana* ARABIDOPSIS THALIANA CENTRORADIALIS (ATC) gene at a coverage of 25% and 74% identity using BLASTn. No protein alignment was found using BLASTx with the SNP sequence, but when the contig it lies within was BLASTed using BLASTp, the same genes were identified. Both genes encode PEBP domain containing proteins and are part of the FT/TFL family. Upon further investigation, the SNP was located to intron 1 of the gene, which could affect the expression of the gene in rocket, leading to the late bolting phenotype.

Once the *De novo* assembly had been built, the six flowering pathway genes identified in rocket during this project were used to search the sequence. Contigs containing these genes were identified, but no SNPs were contained in these regions.

## 5.4. Discussion

An EMS population in rocket was created by treating 4000 *D. tenuifolia* seeds with two concentrations of EMS. At the M<sub>2</sub> generation, late bolting plants were selected and backcrossed to wild type. These were selfed and the late bolting phenotype selected. At the end of the project, two late bolting lines which originated from the same M<sub>2</sub> plant were chosen for genome sequencing analysis. These lines were still segregating into wild type-like bolting and late bolting plants. DNA from each bolting phenotype was pooled and whole genome sequencing performed to identify SNPs which could be linked to the difference in bolting phenotype. The process to create the EMS population through to the SNP identification took the entire life time of the project of four years.

The number of seeds treated with EMS at the beginning was limited to 4000, due to the difficulty of producing enough seed by selfing to ensure a pure line. The small number of seeds which were treated reduced the likelihood of finding individuals with the desired phenotype. The difficulty in producing selfed seed was a problem throughout the generations and many potentially interesting lines were lost due to a lack of selfed seed production. Therefore this continued to hamper the selection of late bolting lines in the population. One backcross to wild type was performed on the M<sub>2</sub> generation late bolting plants to reduce the number of non-causative background SNPs and the late bolting phenotype selected for in the progeny. It would have been useful to perform a second, and possibly a third backcross to wild type on the late bolting lines to further reduce the number of background mutations by 50% per backcross (Stephenson *et al.*, 2010) but this wasn't possible. This should therefore be done on the late bolting lines identified for whole genome sequencing to observe which of the SNPs identified continue to co-segregate with the late bolting phenotype.

The late bolting phenotype selected for in this project was not very strong, and segregation was still apparent in the M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub> generation. Therefore the whole genome sequencing strategy using wild type-like bolting plants against late bolting plants of the same lineage was used in a similar way to work done by Abe *et al.* (2012) (Figure 5.1). This allowed SNPs to be identified which were caused by the EMS treatment which varied between the two samples according to the bolting phenotype. However, the results may have been clearer if the EMS treated lines were taken to the M<sub>3</sub> or M<sub>4</sub> generation before backcrossing to wild type, to increase the number of homozygous alleles in the late bolting lines (Thurling and Depittayanan, 1992). This may also have helped in reducing the pleiotropic mutations of small size, mottled and deformed leaves which segregated along with the late bolting in some plant lines. Natural variation was also present in the seeds used in the EMS treatment. This was because the SSD4 seed used for the EMS treatment was not as pure and uniform as expected, so variation in bolting time was already present in the population before the EMS treatment was used. This has made selection of late bolting more difficult as there was a range of bolting time in the wild type controls as well as a range of bolting times in the EMS population. This may have meant that some individuals scored as late may not have been late due to the treatment, but by the natural variation, while some individuals which were not scored

as late may have been made later by the EMS treatment, but were masked by the natural variation in the population.

Whole genome sequencing of the late bolting and wild type bolting plants of the segregating population revealed five high confidence SNPs with a rank of 1. When these were BLASTed, three showed homology to known genes. SNP 1 aligned to several genes including an ABA sensing gene, SNP 2 to a gene involved in the GA pathway, which suggests that these SNPs might be linked to the late bolting phenotype. The resistance gene SNP 3 is homologous to is unlikely to be causing the late bolting phenotype but it could be genetically linked to the causative SNP due to co-segregation. Of all the high confidence SNPs of ranking 0.7 and above, the one most likely to be linked to the late bolting phenotype was shown to be homologous to an *FT*-like gene, *CENTRORADIALIS*. The *FT* family of genes is known to be involved in floral initiation, so a SNP present in one of these could cause a change in the flowering phenotype. With further investigation, the SNP was found to lie in intron 1 of the rocket *FT*-like gene. SNPs in introns are known to be able to affect gene expression, or splicing, as has been seen in *Platanus acerifolia* (Zhang *et al.*, 2011). Unfortunately, the results from the whole genome sequencing arrived too close to the end of the project for any further analysis, such as analysis of expression of this gene, to be undertaken.

Late bolting individuals in the M<sub>3</sub> and M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub> generations were taken to Elsoms Seeds Ltd for further breeding and crossing into commercial varieties to produce robust late bolting populations.

## Chapter 6. Targeted Mutagenesis

### 6.1. Introduction

The work in this chapter is aiming to explore whether introducing a mutation into specific gene will cause a change in the bolting phenotype of the plant. Six genes in the flowering time pathway have been isolated in rocket and sequenced. Experiments have been carried out to verify the function and behaviour of each gene and compared to Arabidopsis (Chapter 3). With this understanding, it has been possible to choose a candidate gene in the flowering time pathway of rocket which if mutated, would be likely to produce late bolting plants. Producing plants with a late bolting phenotype resulting from the targeted mutagenesis of a known gene would aid our understanding of the role of this gene in the control of flowering in rocket, and would also provide breeding material for new commercial varieties of rocket with delayed bolting. The method of mutagenesis would need to be carefully considered, however, due to the legislation about genetically modified organisms as food sources across the world.

The *FT* gene is a floral integrator gene which needs to be induced for flowering to occur. If a mutation was introduced into the *FT* gene, it would not be able to function to promote flowering thus delaying the floral transition. *FT* has been successfully isolated in rocket (section 3.3.5.1) and is an ideal candidate for targeted mutagenesis. The high homology (99%) between the *DtFT* and *AtFT* genes, has allowed the constructs designed for a clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated system (CRISPR/Cas9) approach to target *FT* in both species. This has been an important feature because it is currently unknown whether rocket is transformable, so the techniques can be applied to Arabidopsis where transformation is known to work and therefore test the functionality of the *FT* targeted CRISPR/Cas9 system if rocket transformation proves unsuccessful.

The use of targeted mutagenesis has enabled scientists to make specific improvements to crops in a much shorter time frame than is possible through conventional breeding. Over the last 20 years, the technology has rapidly evolved. Along with this, there has been much public concern over the use of this technology as the products produced are classed as genetically modified organisms (GMOs).

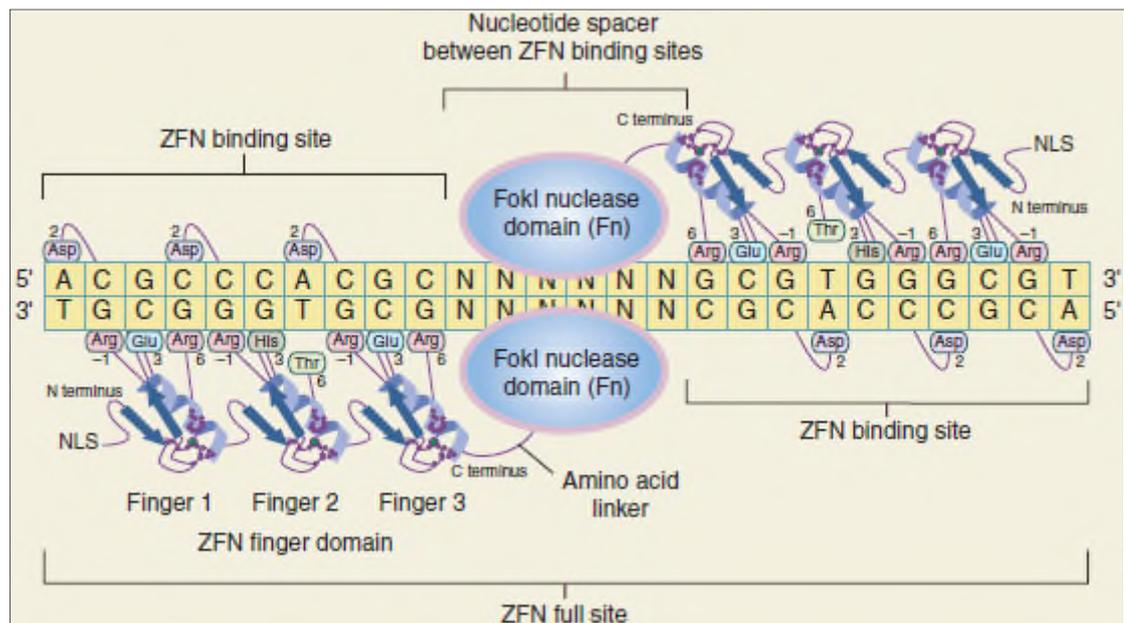
The UK and EU have tight regulations about GMOs and currently do not allow these products to be available for human consumption. However, a paper was written recently which expressed the opinion that products created through the use of targeted mutagenesis (specifically zinc finger nucleases (ZFNs) but also encompassing transcription activator-like effector nucleases (TALENs) and CRISPR/Cas9) pose no more of a hazard to the environment and human health than the techniques currently being used in conventional breeding or gene transfer methods. In fact, the target specificity allowed by targeted mutagenesis could minimise these hazards (EFSA Panel on Genetically modified organisms (GMO) 2012). The USA allow the creation and production of GMOs and the USDA Animal and Plant Health Inspection Service (APHIS) are of the opinion that “*Genetically Engineered plants containing targeted deletions caused by naturally-occurring DNA repair after the targeted break is made by zinc-finger nucleases, and in which no genetic material is inserted into the plant genome, are not regulated articles*” (<http://www.aphis.usda.gov/wps/portal/aphis/home>). This has allowed the development of crops modified using ZFNs to be pursued knowing that they will be able to be brought to market in the USA. It is hoped that this opinion will cover the use of TALENs and CRISPR/Cas9 in the development of GMOs as the mutation introduction by naturally occurring DNA repair is the same. Therefore the use of a targeted mutagenesis method, such as CRISPR/Cas9 in this work to be utilised in a crop species is a viable breeding method which could lead to an improved variety in a shorter time frame than by conventional breeding, with the ability to sell the products in the USA and potentially in the near future, in Europe as well.

#### **6.1.1. Targeted mutagenesis methods**

Targeted mutagenesis is a method which allows a change to occur in the genome of an organism at a specific site. The ability to do this provides a huge potential to crop breeders to quickly improve varieties, by targeting genes involved in quality, disease resistance or drought tolerance amongst many others. Over the last 20 years, the technology has evolved and changed, providing scientists with three methods by which to introduce mutations into a target site within a gene of interest (Bortesi and Fischer, 2015).

#### 6.1.1.1. ZFNs

The first method to be discovered was zinc finger nucleases (ZFNs). The zinc fingers are made up of about 30 amino acids which are linked to a zinc ion and are able to recognise a 3 nucleotide target (Pavletich and Pabo, 1991). To increase the specificity, often 3 or 4 zinc fingers are used together, giving a target region of 9-12 nucleotides. Each codon however, must target ANN, GNN, TNN or CNN, so the number of available target sites within a genome is reduced (Durai *et al.*, 2005). Each set of zinc fingers is associated with a *FokI* restriction endonuclease (from *Flavobacterium okeanokoites*), which enables the target DNA to be cut (Sugisaki and Kanazawa, 1981) and a nuclear localisation signal (NLS) to target the ZFN to the nucleus where it functions. For a mutation to be introduced, a double strand break (DSB) of the DNA must be achieved, and the *FokI* enzyme is only able to do this when it dimerises (Bitinaite *et al.*, 1998). Therefore, to produce a functional ZFN, two sets of zinc fingers each linked to a *FokI* enzyme must be targeted to the sense and antisense strands of DNA, either side of the place where the DSB is to be made (Figure 6.1). This technique has been used successfully in the modification of model and crop plant species (Marton *et al.*, 2010; Shukla *et al.*, 2009; Townsend *et al.*, 2009) but is restricted by the requirements of the recognition codons needed and the introduction of two sets of zinc fingers with *FokI* enzyme to cause the DSB. The technology has been commercialised and software to help design suitable ZFNs to target regions, and companies involved in construction of custom ZFNs are available but are very costly and time consuming (Carroll *et al.*, 2006; DeFrancesco, 2011).

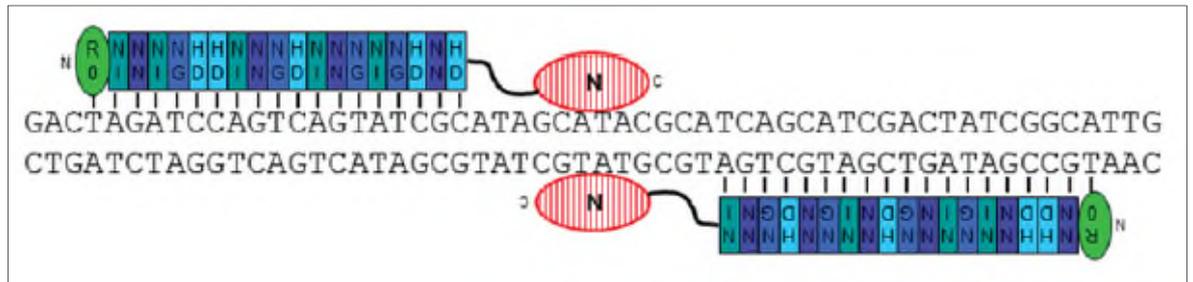


**Figure 6.1 – Diagram of a ZFN consisting of 2 sets of zinc fingers with the dimerised *FokI* endonuclease.** (Porteus and Carroll, 2005). NLS - nuclear localisation signal.

### 6.1.1.2. TALENs

The second method, discovered in 2007 (Roemer *et al.*, 2007), involves transcription activator-like effector nucleases (TALENs). These have been adapted from *Xanthomonas* spp. virulence factors AvrBs3, which are released into the plant to cause disease (Swarup *et al.*, 1992). The TALENs are made up of 1.5-33.5 tandem repeats, and each repeat consists of 33-35 amino acids. The amino acids at positions 12 and 13 are able to recognise and therefore target a single nucleotide (Deng *et al.*, 2012). A TALEN consisting of at least 11.5 tandem repeats (a 12-bp target site) gives a strong binding to the target site, but the greater the number of repeats, the better the target specificity (Morbitzer *et al.*, 2010). TALENs also contain a NLS, so are able to enter into the nucleus and bind with the DNA, and as with ZFNs, are linked to a *FokI* restriction endonuclease to confer the DNA cutting ability. TALENs therefore, must also be used in pairs, so the *FokI* can dimerise to produce a DSB in the target DNA (Mahfouz *et al.*, 2011) (Figure 6.2). TALENs can be designed much more easily than ZFNs as each repeat recognises one nucleotide, so repeats specific to the desired nucleotides can be joined together to target a specific site in a gene of interest. However, as with ZFNs, two TALENs must be introduced in to the target organism for the mutation to be caused (Cermak *et al.*, 2011). TALENs have been used successfully to introduce mutations into plants, in both

monocot and dicot species (Christian *et al.*, 2013; Shan *et al.*, 2013a; Zhang *et al.*, 2013). TALEN technology has also been commercialised with companies able to provide design and production services at a much lower cost and timeframes than for ZFNs (Cermak *et al.*, 2011; DeFrancesco, 2011; Reyon *et al.*, 2012; Xiao *et al.*, 2013).

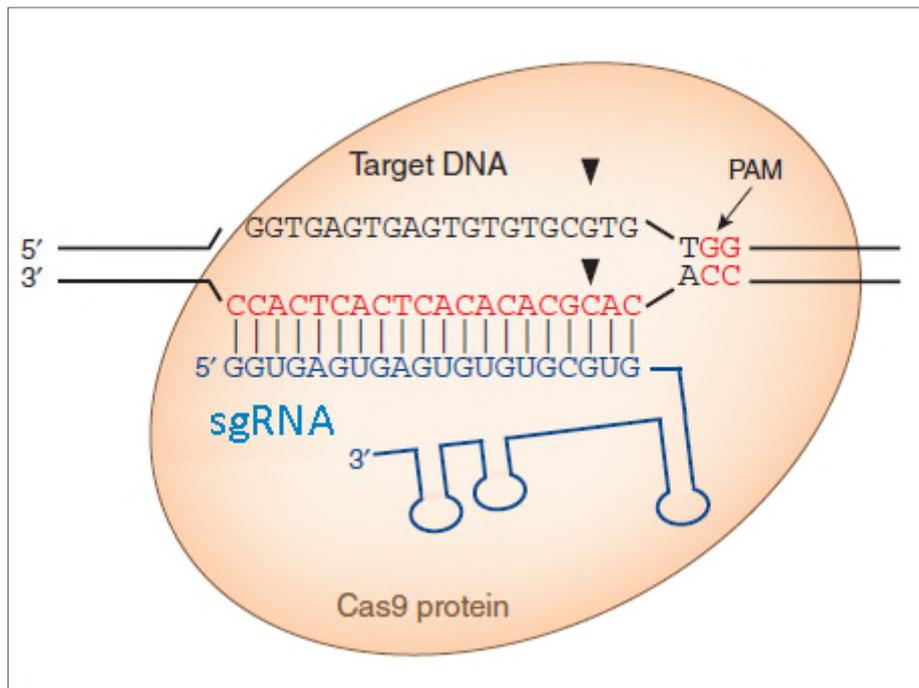


**Figure 6.2 – A pair of TALENs showing orientation and function when aligned to a target gene.**  
N - *FokI* nuclease. (Wright *et al.*, 2014)

### 6.1.1.3. CRISPR/Cas9

The third method of targeted mutagenesis, discovered in 2012, is the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated system (cas9). CRISPR/Cas9 is found in bacteria and archaea as a defence mechanism against invading foreign DNA (Jinek *et al.*, 2012). The organism uses the incoming DNA to produce target RNA linked to the Cas9 protein, which is then able to cut and disable any further exogenous DNA of the same type. Scientists have utilised the CRISPR/Cas9 system of *Streptococcus pyogenes* to produce an efficient, functional targeted mutagenesis method (Bortesi and Fischer, 2015). The Cas9 protein is an endonuclease capable of causing a DSB, and in *S. pyogenes* (and other bacteria and archaea) is linked with two RNAs, a CRISPR RNA (crRNA) which determines the specificity of the CRISPR, and a trans-activating CRISPR RNA (tracrRNA) which is necessary for stabilisation of the CRISPR/cas9 structure. The target region must be upstream of a protospacer-adjacent motif (PAM) site for the Cas9 to cause a DSB (Hsu *et al.*, 2013). To ease the targeting and construction of the CRISPR/Cas9, the tracrRNA and crRNA have been combined to give one single guide RNA (sgRNA) which is able to combine the specificity and stabilisation of the structure (Jinek *et al.*, 2012). The sgRNA is 120 nucleotides long and includes the 20 nucleotide target region which must be upstream of a PAM site (5'-NGG or less efficient 5'-NAG) for the DSB to occur (Hsu *et al.*, 2013; Pattanayak *et al.*, 2013). Due to the nature of the

Cas9 protein targeting and cutting, the DSB will be made three nucleotides upstream from the PAM, within the sgRNA target region (Mali *et al.*, 2013) (Figure 6.3). Despite its short history, this method has been applied widely, to both mammalian systems (particularly mice, zebrafish and humans (Hsu *et al.*, 2013; Li *et al.*, 2013a; Mali *et al.*, 2013; Wang *et al.*, 2013)), and to plants (both model and crop species (Jiang *et al.*, 2013; Li *et al.*, 2013b; Nekrasov *et al.*, 2013; Shan *et al.*, 2013b)). It has been successful in producing mutations at target sites, but with varied efficiencies (Bortesi and Fischer, 2015). The major advantages of the CRISPR/Cas9 system over ZFNs and TALENS are the speed and low cost of production of the construct, and that only a single molecule needs to be introduced to an organism to cause a DSB, rather than two, so transformation occurs more easily and more efficiently. This is why it is the method of choice for this study. However, one disadvantage for the use of CRISPR/Cas9 is the greater occurrence of off-target mutations (Fu *et al.*, 2013). It has been found that the sgRNA only needs the 8-12 nucleotides nearest the PAM site to be correct and specific for the targeting and DSB to take place, with mismatches in the rest of the 20 nucleotides being tolerated (Pattanayak *et al.*, 2013). This therefore increases the likelihood of the CRISPR/Cas9 being able to bind elsewhere in the genome. However, this is less of a problem in plants as these unwanted mutations can be easily bred out and constructs can be optimised to reduce off-target cuts.

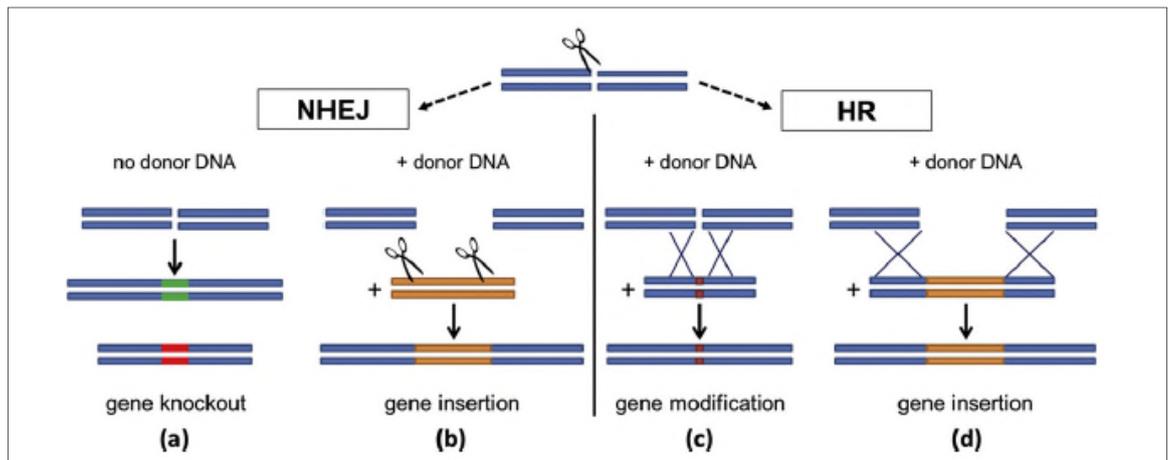


**Figure 6.3 – CRISPR/Cas9 system.**

The Cas9 protein is located to the target DNA by the sgRNA, adjacent to the PAM. The DSB is made three nucleotides upstream from the PAM site (Black triangles). The 3' end loop structures of sgRNA aid the specificity and stabilisation of the CRISPR/Cas9 (Carroll, 2013).

#### **6.1.1.4. DNA repair mechanisms**

All three methods rely on errors made by the naturally occurring DNA repair systems, non-homologous end joining (NHEJ) or homologous recombination (HR) to go from a DSB to a mutation by the insertion or deletion (InDel) of nucleotides. NHEJ tends to be the more frequently used method of repair as it doesn't require any additional information, but it means that the InDel is random. However, both can be utilised by the researcher to introduce a specific sequence of DNA into the target region by transforming in a piece of donor DNA together with the CRISPR/Cas9. In NHEJ, the donor DNA is also cut by the Cas9 nuclease and joined into the gene sequence at the DSB site. In HR, the donor DNA acts as a template for the DSB site to be repaired by nucleotide substitution (Figure 6.4).



**Figure 6.4 – Comparison of NHEJ (a+b) and HR (c+d) methods of DNA repair.**

a) Repair mechanism causes insertion (green) or deletion (red) of nucleotides. b) If donor DNA carrying the CRISPR target sequence at each end is added with the CRISPR/Cas9 construct, this too will be cut and then inserted into the gene. c) The addition of donor DNA carrying a single nucleotide at the site of mutation, when the target DNA is cut, this will be used as a template in the DNA repair and so incorporated. d) Where larger insertions are desired, a donor DNA carrying this between sequences homologous to those on either side of the cut site, the matching DNA is lined up after the cut has been made and used as a template as the repair occurs (Bortesi and Fischer, 2015).

## 6.1.2. Optimisation of CRISPR/Cas9 methods

### 6.1.2.1. Construction of plant specific CRISPR/Cas9

The CRISPR/Cas9 system from *S. pyogenes* has been adapted for targeted mutagenesis in humans (Cong *et al.*, 2013b), but has also needed to be adapted for use in plants. Work done by Feng *et al.* (2013) optimised the Cas9 construct so it would be functional in plants, by replacing the promoters with two 35S promoters from Cauliflower Mosaic Virus (CaMV) and adding a Nos terminator. The CRISPR sgRNA also needed a plant specific promoter, so they used the Arabidopsis AtU6-26 promoter and fused this to the 85bp chimeric guide RNA sequence which contains restriction enzyme sites to allow the insertion of the specific 20 nucleotide CRISPR target sequence. It was this research group that provided the Cas9 and AtU6-26 constructs for use in this project.

### 6.1.2.2. Delivery of CRISPR/Cas9 into plant cells

Many plant species, including *A. thaliana*, *Nicotiana benthamiana*, *Oryza sativa* and *Triticum aestivum*, have been successfully gene-edited using CRISPR/Cas9. There has been a variety of methods used to introduce the CRISPR/Cas9 construct into the plant (Bortesi and Fischer, 2015). Leaf agro-infiltration has been used in both

Arabidopsis and *N. benthamiana* and uses *A. tumefaciens* to deliver the CRISPR/Cas9 plasmid into the plant cells. It is a commonly used method for transformation, and has shown success with CRISPR/Cas9, but with a mutation frequency of 2.7-13.8% (Jiang *et al.*, 2013; Li *et al.*, 2013b; Nekrasov *et al.*, 2013; Upadhyay *et al.*, 2013). PEG-protoplast transfection is also a commonly used technique and has been employed with more success than agro-infiltration. The mutation frequency has varied from 1.1-38.5% and used in Arabidopsis and *N. benthamiana* (Li *et al.*, 2013b), rice (Jiang *et al.*, 2013; Shan *et al.*, 2013b), wheat (Wang *et al.*, 2014) and maize (Liang *et al.*, 2014). However, this method is dependent on having species which are amenable to protoplast isolation and transformation. Agro-transformation using the floral dip method has been used repeatedly in Arabidopsis and gives a wide range of mutation frequencies, from 0.04-80%, depending on the exact CRISPR/Cas9 used and the genes it is targeted to (Fauser *et al.*, 2014; Feng *et al.*, 2013; Mao *et al.*, 2013). Agro-transformation of callus has been used as a viable transformation method in rice, with mutation frequencies ranging from 0-100%, but with most >50%. This is again dependent on the CRISPR/Cas9 construct used and genes targeted (Feng *et al.*, 2013; Mao *et al.*, 2013; Miao *et al.*, 2013). For this study, protoplast isolation and transformation, and agro-transformation by floral dipping will be the methods of choice due to the high mutation frequencies seen in the literature and the amenability of Arabidopsis to both methods. Attempts will be made to transform rocket plants as well using both methods; floral dipping which hasn't been tried before and protoplast isolation which has been achieved (Chatterjee *et al.*, 1988; Sikdar *et al.*, 1990), although transformation of rocket protoplasts has not been reported.

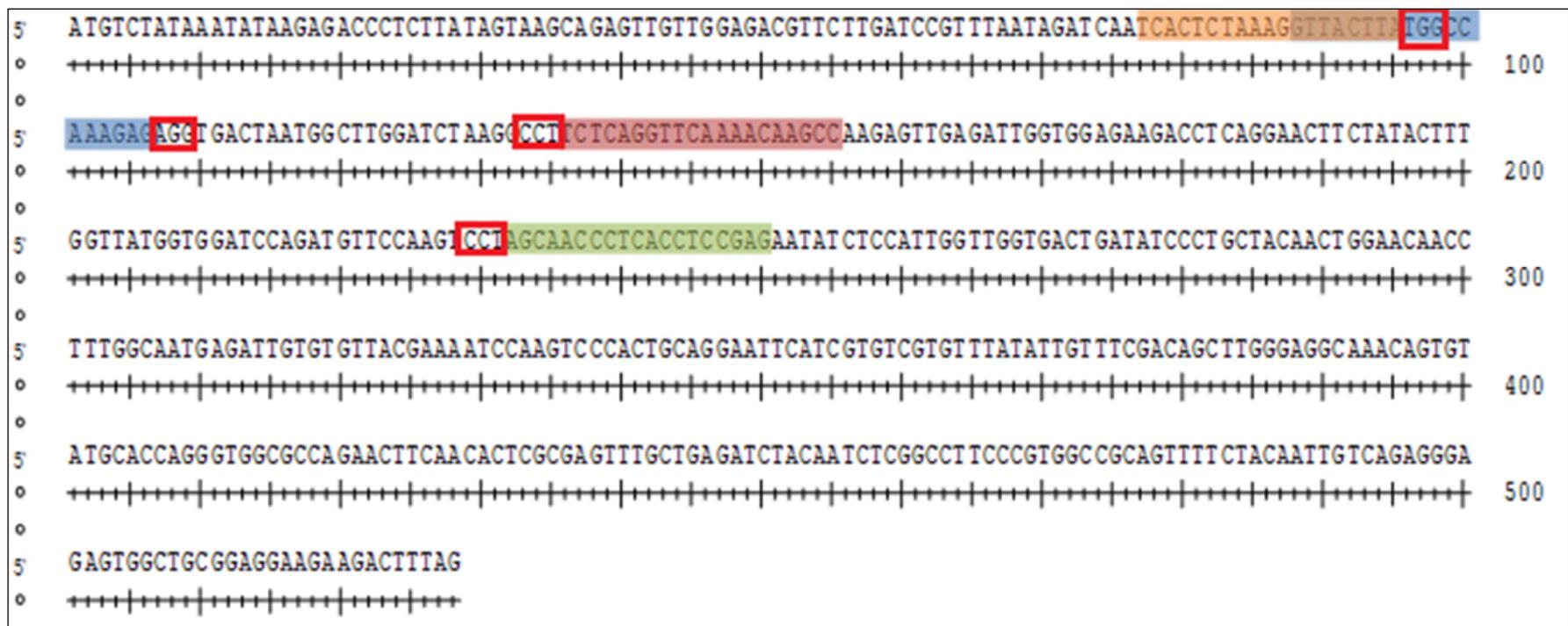
## **6.2. Methods**

### **6.2.1. CRISPR design and construction**

#### **6.2.1.1. CRISPR target site sequence design**

The coding sequence for *FLOWERING LOCUS T (FT)* from rocket (*DtFT*) has been isolated (section 3.3.5.1) and was used to design the CRISPR target sequences. The *DtFT* gene coding sequence has only one base difference to the Arabidopsis *FT* (*AtFT*) gene so it is expected that the CRISPRs that are designed to *DtFT* sequences not including this SNP will also work on *AtFT*. Four regions of the gene were

chosen to design the CRISPR target sequences to, to increase the likelihood that one may work efficiently and cause the late bolting phenotype (Figure 6.5). Feng et al. (2013) and Jiang et al. (2013) provided guidance on how to design the CRISPRs which included the requirement for a PAM site (NGG) to be adjacent to the CRISPR target sequence for the system to be functional. Constructs FT-1 and FT-2 were designed by analysing the *DtFT* sequence for a suitable site, and constructs FT-96 and FT-136 were designed using E-CRISP ([www.e-crisp.org](http://www.e-crisp.org)) and MIT ([www.crispr.mit.edu](http://www.crispr.mit.edu)) methods. At one end of the CRISPR target sequence, a *BbsI* restriction site and a G nucleotide must be included in the primer to allow the subsequent reactions for the construction to take place. Each CRISPR target sequence also contains a restriction site for post-mutagenesis analysis. The CRISPR target sequences were ordered as primers (Invitrogen) with forward and backward sequences covering the same region so that they could be annealed together to create a double stranded piece of DNA for the construct (Appendix D.2.1).

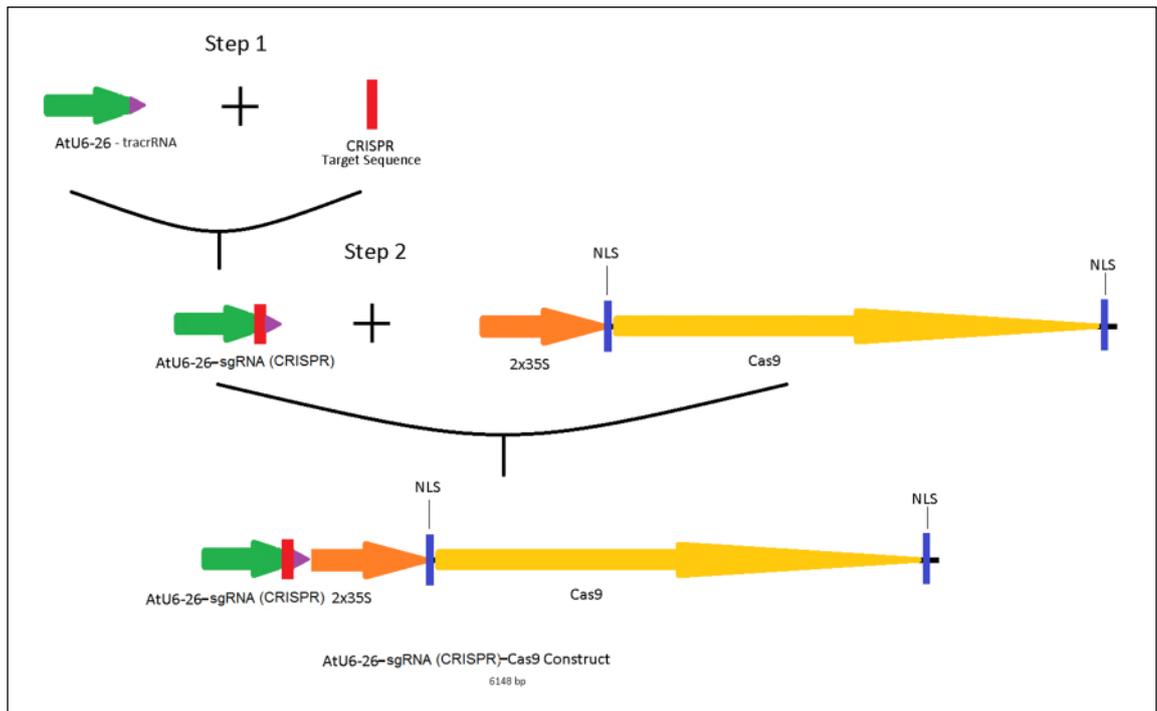


**Figure 6.5 – DtFT cds with sequences targeted by the CRISPR constructs highlighted.**

Four CRISPR constructs were designed to different sequences in the DtFT cds. Orange – FT-96, Blue – FT-1 (brown depicts the region of overlap of FT-96 and FT-1 sequences), Red – FT-136, Green – FT-2. PAM sites are shown by red boxes.

#### **6.2.1.2. Creating the CRISPR-Cas9 constructs (carried out at Hangzhou Normal University under guidance from Dr Zhiming Yu)**

Primer pairs for each CRISPR target sequence were annealed together, and ligated into a pBluescript SK+ plasmid containing the AtU6-26 promoter and sgRNA provided by Dr. Botao Zhang, Chinese Academy of Sciences, Shanghai (Feng *et al.*, 2013) which had been cut at the *BbsI* restriction site (Figure 6.6-Step1). The plasmid was transformed into *E.coli* and colony PCR was performed using M13 primers and CRISPR specific primers to test for the presence of the *FT* CRISPR target sequence in the AtU6-26+sgRNA construct. Positive clones were mini-prepped (section 2.2.8.2) and digested with *SalI* and *KpnI* restriction enzymes. To make the transient constructs in pBluescript SK+, the Cas9 protein with its 35S promoter DNA (adapted for use in Arabidopsis and received from Dr. Botao Zhang, Chinese Academy of Sciences, Shanghai) (Feng *et al.*, 2013) was also digested with *SalI* and *KpnI* restriction enzymes. This was ligated to the AtU6-26-sgRNA CRISPR unit (Figure 6.6-Step2) and transformed into *E.coli*. Colonies were picked and miniprepped. Digestion reactions were set up using *SalI* and *KpnI* restriction enzymes to check for the presence of all construct components in pBluescript SK+ (AtU6-26-sgRNA+CRISPR+35S-Cas9). To make the stable constructs in pCAMBIA1300, the Cas9 protein with its 35S promoter DNA was digested with *SalI* and *EcoRI* restriction enzymes and the pCAMBIA1300 vector digested with *KpnI* and *EcoRI* restriction enzymes. These digested products along with the AtU6-26-sgRNA+CRISPR target sequence digested product were ligated together and transformed into *E.coli* (Figure 6.6-Step2). Colonies were picked, miniprepped and digested using *KpnI* and *EcoRI* to check for the presence of all construct components in pCAMBIA1300 (AtU6-26-sgRNA+CRISPR+35S-Cas9 (Figure 6.6)). Each of the four CRISPR constructs in the pCAMBIA 1300 or pBluescript SK+ vector was transformed into EC100 cells via electroporation. This was then plated onto LB (10:5:10)/1.5% (w/v) agar plus the antibiotic and incubated at 37°C for 16-18 hours. Clones were picked and used to inoculate 8ml LB (10:5:10) plus antibiotic before miniprepping using a Plasmid DNA Extraction kit (Thermo Scientific) to extract the CRISPR construct plasmid DNA (section 2.2.8.2) and sequencing the DNA.



**Figure 6.6 – Complete construction of AtU6-26-CRISPR-Cas9.**

Step 1 – AtU6-26 promoter + tracrRNA is ligated to double stranded *FT*-CRISPR target sequence. Step 2 – AtU6-26-sgRNA+CRISPR unit is ligated to Cas9 and Cas9 promoter unit to give the complete CRISPR/Cas9 construct. This part of the construct is the same for both transient and stable transformation approaches, but the vector backbone is different. NLS – nuclear localisation signal

## 6.2.2. Stable transformation

### 6.2.2.1. Floral dip transformation of *Arabidopsis* and *D. tenuifolia*

The CRISPR construct DNA was used to transform *A. tumefaciens* (strain c58pGV3101) cells (section 2.2.9). A large volume solution of the transformed cells was produced and used for floral dipping of *Arabidopsis* Col0 (Clough and Bent, 1998) and *D. tenuifolia* (rocket) SSD6 plants. *Arabidopsis* plants were allowed to self and set seeds, but the rocket plants were bud pollinated to ensure self-seeds were produced (section 2.1.4.2). Seeds were collected and threshed (section 2.1.5).

### 6.2.2.2. Selection of CRISPR transformed plants

Seeds were sterilised by placing into glycine packets and arranging in a plastic box so packets don't overlap. 100 ml 10% Sodium hypochlorite solution was put in a beaker and 3 ml 32% hydrochloric acid poured on top to release the chlorine gas. The plastic box was sealed and left in a fume hood for 16-18 hours. Seed packets were removed and aerated for 20 minutes in a sterile flow cabinet before plating. *Arabidopsis* Col0 and rocket wild type seeds were mixed with 0.1% sterile agarose

solution and spread evenly onto plates made of 2.2 g/l MS, 30g/l sucrose and 7 g/l agar (pH5.7-5.8). Wild type and CRISPR-transformed Arabidopsis Col0 and rocket seeds were mixed with 0.1% sterile agarose containing 100 µg/ml Cefataxime and spread evenly onto plates made of 2.2 g/l MS, 30g/l sucrose and 7 g/l agar (pH5.7-5.8) containing 20 µg/ml Hygromycin B. All plates were stratified at 4°C for three days. These were then kept in the dark at 22°C for germination and hygromycin selection to take place (about seven days). The plates were then put into a 16 hour photoperiod for 3-4 days until the plants were big enough to be transplanted onto soil. Plants which had survived the selection and were large and green compared to the rest, were picked off the plate and transplanted into Arabidopsis mix soil (Levington F2s : sand : vermiculite fine grade 6:1:1) at one per cell in p24 trays and placed into Panasonic controlled environment cabinets under 22°C, 16 hour light, 8 hour dark conditions. Rocket plants were transplanted a second time into M2 soil in 5 inch pots at 5-6 leaves. Bolting was measured by rosette leaf number at 1 cm bolt for Arabidopsis transformants and leaf number at floral initiation for rocket transformants.

#### **6.2.2.3. Identification of InDels caused by CRISPR constructs**

Leaf material was collected from all CRISPR transformed plants and DNA extracted using the CTAB method (Stewart and Via, 1993) (section 2.2.1). For Arabidopsis lines, the full length *FT* gene was amplified using *AtFT\_5'UTR\_1F* and *AtFT\_3'UTR\_1R* primers by PCR and the product purified using a Qiagen PCR purification kit (section 2.2.5). The products were then sequenced using *AtFT* specific primers and compared to the wild type (WT) Arabidopsis *FT* gene sequence (Appendix D.2.2). This process was not carried out for rocket transformants as the full length *DtFT* gDNA gene has not currently been isolated. For further confirmation of InDels causing the late bolting phenotype the CRISPR target sequences were cloned. PCR was used to amplify shorter sequences from *AtFT* to cover the regions to which CRISPR FT-1 and FT-2 had been designed to (Figure 6.5 and Appendix D.2.3). These were A-tailed and products were PCR purified using a Qiagen PCR purification kit and eluted in 20 µl before quantifying (section 2.2.5). Each was then ligated into pGEM-T easy vector system. Five clones from each transformation were picked and plasmid DNA extracted (section 2.2.8.2) and

sequenced using M13 primers and compared with WT *AtFT* and CRISPR FT-1 or FT-2 sequence to see if any InDels were present.

### **6.2.3. Transient transformation**

#### **6.2.3.1. Protoplast isolation**

Two methods for isolating protoplasts were used. Method 1 had been optimised for use with *Arabidopsis* leaf material (Yoo *et al.*, 2007), but was used in these experiments for both *Arabidopsis* and rocket. 3-4 week old plants were used and the newest, fully expanded leaves were picked. Leaf tips and petioles were cut away and the leaves finely sliced. These were placed into freshly made enzyme solution containing 20 mM MES (pH 5.7), 1.5% (w/v) cellulose Onozuka R10 (Duchefa), 0.4% (w/v) maceroenzyme R10 (Duchefa), 0.4 M mannitol and 20 mM KCl (warmed to 42°C), and submerged. A vacuum was applied to draw the enzyme solution into the air spaces in the tissue and leaves placed on a tilting platform for 1-2 hours. An equal volume of cold W5 solution (2 mM MES (pH 5.7), 154 mM NaCl, 125 mM CaCl<sub>2</sub> and 5 mM KCl) was added and this was filtered through a 70 µm sieve. The protoplast solution was centrifuged at 100 g for 3 min at 4°C and the supernatant removed carefully. Protoplasts were resuspended in 2 ml MMG solution (4 mM MES (pH 5.7), 0.4 M mannitol and 15 mM MgCl<sub>2</sub>) and kept on ice.

The second method had been optimised for *Eruca sativa* and *D. muralis* (Chatterjee *et al.*, 1988; Sikdar *et al.*, 1987; Sikdar *et al.*, 1990) and was used for both *Arabidopsis* and rocket leaf material. The leaf material was harvested and sliced as described above, but the enzyme solution contained 0.6 M mannitol, 0.2% (w/v) CaCl<sub>2</sub>.H<sub>2</sub>O, 0.25% (w/v) cellulose Onozuka R-10 (Duchefa) and 0.5% (w/v) pectinase (Sigma) at pH 5.8. The leaf pieces were submerged and the plate put at 28°C in the dark for 16-18 hours. The cell solution was washed through a 70 µm sieve using 20 ml of 20% (w/v) sucrose solution and centrifuged at 100 g for 3 min at 4°C. After the supernatant was removed, 10 ml Sikdar's osmoticum (0.6 M mannitol and 0.2% (w/v) CaCl<sub>2</sub>.H<sub>2</sub>O) was used to resuspend the cells and centrifuged at 100 g for 3 min at 4°C. The supernatant was removed and another 10 ml Sikdar's osmoticum added and the centrifugation repeated. Once the supernatant had been removed, the protoplasts were finally resuspended in 5 ml Sikdar's osmoticum and kept on ice.

Following protoplast isolation using either method, the protoplast suspension was put onto a Fuchs-Rosenthal haemocytometer and all non-disrupted cells counted under a light microscope (Olympus BH-2) and the total number of cells per ml calculated. 100  $\mu$ l cells were mixed with 0.5  $\mu$ l (5 mg/ml) Fluorescein Diacetate (FDA) and covered for 10 mins. The cell suspension plus FDA was then applied to the Fuchs-Rosenthal haemocytometer and fluorescing cells counted using a fluorescent microscope (Nikon Optiphot) to estimate the number of viable cells per ml.

### **6.2.3.2. Protoplast transformation**

A 5 ml cell culture was used to inoculate 500 ml LB (10:5:10) plus 100  $\mu$ g/ml Ampicillin and incubated at 37°C for 16-18 hours. A Plasmid Maxi kit (Qiagen) was used to extract the plasmid DNA from the cell culture in large enough quantities for the protoplast transformation protocol to be used several times according to the manufacturer's instructions. Plasmid maxi-preps were also used to prepare sufficient plasmid DNA from the vector pG201 expressing Green Fluorescent protein (GFP), vector pGY-1 expressing Yellow Fluorescent protein (YFP) and pGY-1 expressing mCherry (Weis *et al.*, 2013). Plasmid pG201-GFP was donated by Dr Daniel Tome, University of Warwick and pGY-1-YFP and pGY-1-mCherry plasmids were donated by Dr Ruth Schäfer, University of Warwick.

To perform the protoplast transformation from Method 1 (Yoo *et al.*, 2007), a fresh PEG-calcium solution was made (1 g PEG 4000, 0.625 ml 0.8 M mannitol, 0.25 ml 1 M CaCl<sub>2</sub>, 1.625 ml H<sub>2</sub>O). The protoplast solution was diluted with MMG to reach 200,000 cells/ml, or 400,000 cells/ml depending whether a single or double transformation was to be carried out. The desired CRISPR and fluorescent reporter plasmid DNA was put into a 2 ml tube and protoplasts added on top in the amounts shown in Table 6.1. The PEG-calcium solution was added and the tube flicked until all solutions had combined. The solution was put at room temperature for 30 min before adding 500  $\mu$ l cold W5 to stop the reaction. Cells were centrifuged at 100 g for 3 min and the supernatant removed. 500  $\mu$ l W1 solution (4 mM MES (pH 5.7), 0.5 M mannitol and 20 mM KCl) was used to resuspend the cells, and tubes were covered with breathable Nescofilm and put on a windowsill for 24 hours. Cells were then analysed using a fluorescent microscope (Nikon Optiphot) or a fluorescent activated cell sorting (FACS) machine (BD Influx™ Cell Sorter) depending on whether the reporter plasmid used contained GFP or YFP.

**Table 6.1 – Quantities of protoplast cells, plasmid and PEG-calcium solution used for transformation of protoplast cells.**

Adapted from Yoo *et al.*(2007)

<b>Single transformation</b>	20,000 protoplasts	40,000 protoplasts	60,000 protoplasts
Plasmid	4 µg	8 µg	12 µg
Protoplast solution (200,000 cells/ml)	100 µl	200 µl	300 µl
PEG	100 µl + vol. plasmid	200 µl + vol. plasmid	300 µl + vol. plasmid
<b>Double transformation</b>	40,000 protoplasts	80,000 protoplasts	120,000 protoplasts
Plasmid 1	4 µg	8 µg	12 µg
Plasmid 2	4 µg	8 µg	12 µg
Protoplast solution (400,000 cells/ml)	100 µl	200 µl	300 µl
PEG	100 µl + vol. plasmids	200 µl + vol. plasmids	300 µl + vol. plasmids

### **6.2.3.1. Regeneration of protoplasts to produce plants**

The method for regenerating protoplasts was adapted from Sikdar *et al.* (1990) who had developed this for *D. muralis*. In this experiment, protoplasts from both *Arabidopsis* and *D. tenuifolia* were used as the starting material.

**Table 6.2 – Media recipes needed for regenerating protoplasts**

(Sikdar et al., 1990)

Media Name	Basal Salts used	Sugar content	Hormones	Agar	Media pH
M8PS2	Kao and Michayluk vitamin solution	0.6M glucose	0.2mg/l 2,4-D, 0.5mg/l NAA, 0.5mg/l BAP, 0.5mg/l GA3	None	5.8
M8PS21	Kao and Michayluk vitamin solution	0.4M glucose, 0.0291M sucrose	0.2mg/l 2,4-D, 0.5mg/l NAA, 1mg/l BAP, 0.5mg/l GA3	None	5.8
DP1	Murashige and Skoog	0.0874M sucrose	0.4mg/l 2,4-D, 0.3mg/l NAA, 1mg/l BAP, 0.5mg/l GA3	0.5%	5.8
D1	half strength Murashige and Skoog	0.0436M sucrose	1mg/l NAA, 1mg/l BAP, 0.5mg/l GA3	0.8%	5.8
D13	Murashige and Skoog	0.0874M sucrose	1mg/l BAP, 0.5mg/l GA3	0.8%	5.8
D13-H	Murashige and Skoog	0.0874M sucrose	None	0.8%	5.8

Under sterile conditions, 100,000 cells/ml were pipetted onto a 35 mm petri dish and 2 ml M8PS2 media was added and gently mixed. The plate was put into a Panasonic controlled environment cabinet at 24°C in the dark for 10 days. On day 5, the first cell division should occur. On day 10, protoplasts should be at the 16-64 cell stage and were transferred to 2 ml M8PS21 media. This was placed back at 24°C in the dark for another 10 days. On day 20, protoplasts should have formed a globular or heart shaped embryo and these were carefully transferred to 3 ml DP1 media. Plates were returned to dark conditions but the temperature increased to 25°C. On day 30, greenish nodule microcalli should be visible and excised before transferring to 3 ml D1 media and placed into 25°C, 16 hour photoperiod under low intensity lights conditions for 15 days. On day 45, shoot development should have begun and these

calli transferred onto 55 mm petri dishes containing D13 media and placed under the same conditions as the previous step. At day 60, roots would have begun to initiate and calli transferred to 55 mm petri dishes containing D13-H media and returned to the Panasonic cabinet under the previous step conditions. By day 66-70, the complete plantlet should have formed and when this is clear, the plantlet should be transferred to soil and grown in shade for 7 days and then uncovered.

## **6.3. Results**

### **6.3.1. Analysis of transformed plants**

#### **6.3.1.1. Production of CRISPR constructs**

The double stranded oligos for each of the *FT* CRISPR targets were ligated together with the AtU6-26 and 35S-Cas9 units to produce the CRISPR/Cas9 constructs. Each construct was ligated into pCAMBIA1300 and pBluescript SK+ to enable both stable and transient transformation. Sequencing using M13 and CRISPR-specific primers was carried out to ensure the CRISPR target region DNA was correctly inserted into the construct.

Eight constructs were successfully produced from four CRISPR target regions in two different vectors. Both CRISPR FT-96 constructs were designed and made by Dr Zhiming Yu at Hangzhou Normal University. Figure 6.7 shows an alignment for the four constructs highlighting the CRISPR target regions embedded into the AtU6-26 unit. This sequence was observed in both pCAMBIA1300 and pBluescript constructs. The confirmation of all eight constructs allowed the transformation of plants and protoplasts to be started.

Sequence Name	< Pos = 484											
- ■■■■■ +												
<input checked="" type="checkbox"/> Consensus	TAGAGTCGAAAGTAGTGATTGGXTXGTTATGXGCXXXGXA G T T T TAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT											
4 Sequences	490	500	510	520	530	540	550	560	570	580	590	600
CRISPR FT-1	TAGAGTCGAAAGTAGTGATTGGTTACTTATGGCCAAAGAG G T T T TAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT											
CRISPR FT-2	TAGAGTCGAAAGTAGTGATTGCTCGGAGGTGAGGGTTGCT G T T T TAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT											
CRISPR FT-96	TAGAGTCGAAAGTAGTGATTGCTCACTCTAAAAGTTACTTAG T T T TAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT											
CRISPR FT-136	TAGAGTCGAAAGTAGTGATTGGCTTGT T T T GAACCTGAGA G T T T TAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT											

**Figure 6.7 – Alignment of CRISPR constructs showing insertion of CRISPR target regions.**

Alignment performed on sequences isolated for each CRISPR construct using DNA Lasergene 9 core suite MegAlign showing successful integration of the CRISPR target region DNAs (red box) into the AtU6-26-tracrRNA (unboxed region).

### 6.3.1.2. Selection of CRISPR construct transformed plants

The CRISPR constructs were transformed into Arabidopsis and rocket using Agrobacterium-mediated floral dipping. Seeds were harvested from the transformed plants and sterilised before sowing onto MS plus hygromycin plates for the selection of transformed seeds. This process was carried out twice to ensure enough transformed plants were recovered. The number of surviving plants was recorded (Table 6.3). Experiment 1 provided good numbers of transformed Arabidopsis plants for constructs FT-1, FT-2 and FT-96. Arabidopsis FT-136 seeds were not sown out in this experiment. Six transformed rocket plants were recovered in experiment 1, one with the FT-2 construct and five with the FT-96 construct. Rocket FT-136 seeds were not sown out in this experiment. Experiment 2 was more successful for Arabidopsis transformed plants as a higher number of plants survived. Arabidopsis FT-96 seeds were not sown out in this experiment. However, no transformed rocket plants survived the hygromycin selection.

**Table 6.3 – Number of transformed plants surviving hygromycin selection for each experiment.**

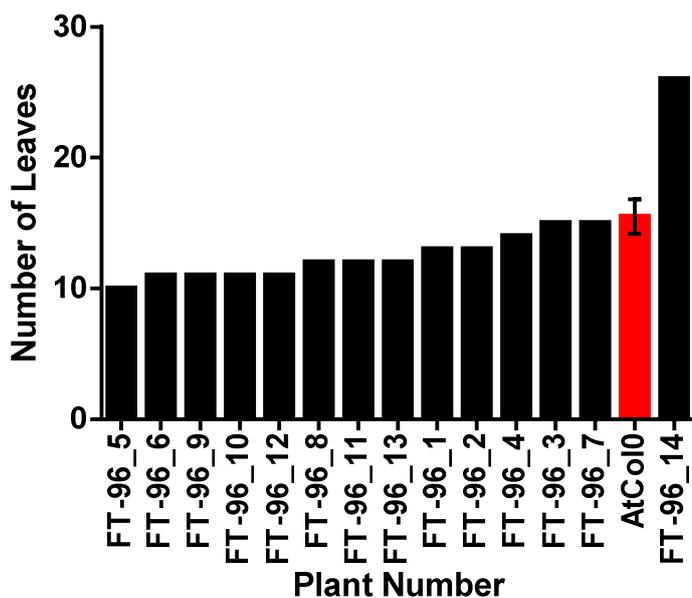
Arabidopsis and rocket plants were transformed with each of four stable CRISPR/Cas9 constructs. Seeds were harvested and hygromycin used to select for transformed seeds. The number of surviving plants for each construct and plant background was recorded. n/a – these transformed seeds were not sown out in this experiment.

		Number of surviving transformed plants	
Plant	Construct	Experiment 1	Experiment 2
Arabidopsis	FT-1	14	16
Arabidopsis	FT-2	11	24
Arabidopsis	FT-96	14	n/a
Arabidopsis	FT-136	n/a	13
Rocket	FT-1	0	0
Rocket	FT-2	1	0
Rocket	FT-96	5	0
Rocket	FT-136	n/a	0

### **6.3.1.3. Bolting phenotypes of *At*:CRISPR and *Dt*:CRISPR transformed plants**

The seedlings which survived the hygromycin selection in each experiment were transplanted onto Arabidopsis mix soil and grown in a controlled environment cabinet at 22°C under long day conditions. The flowering time in LD of the transformed Arabidopsis plants was measured using rosette leaf number (no more rosette leaves are formed after a plant starts to flower). *AtCol0* was used as the wild type control and the bolting phenotype of these plants varied between experiments, in experiment 1, *AtCol0* started flowering at 42 days and 15.5 leaves (average of four plants), and was the wild type control for the transformed lines FT-96, FT-1 and FT-2. In experiment 2, *AtCol0* flowered at 33 days and 10.5 leaves (average of four plants) and was the wild type control for FT-1, FT-2 and FT-136. The discrepancy between the two values could have been due to the experimental conditions, where the timing of transplanting varied between experiment 1 (three days post germination) and experiment 2 (twelve days post germination), or the fungal contamination of the agar plates that occurred in experiment 1 that may have stressed the seedlings, or that covering transplanted seedlings with propagator lids in experiment 2 for a few days could have sped up development due to warmer growing conditions. These experiments comprise of a screen, therefore statistical analysis could not be performed on the data due to the small number of plants in each experiment. When mutated plants are identified and selfed, the T<sub>2</sub> generation is expected to segregate with a 3:1 ratio where valid statistical analysis can be performed.

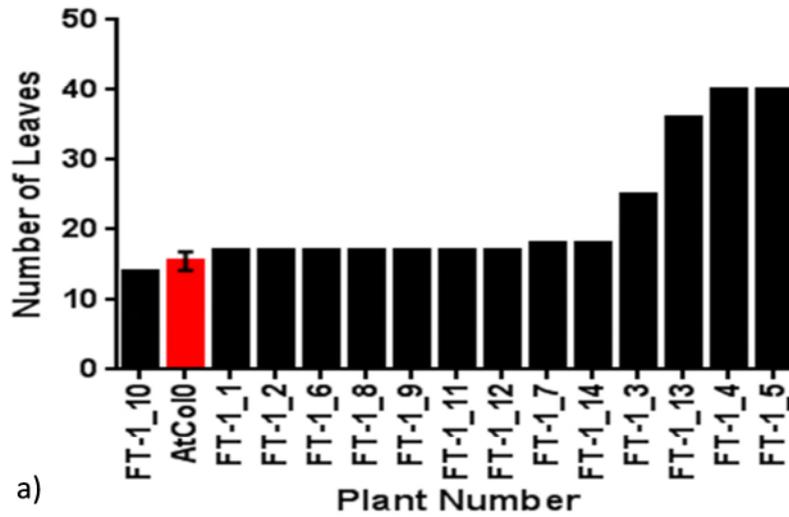
The flowering time results for CRISPR construct FT-96 transformed T<sub>1</sub> Arabidopsis plants show that only one plant was later than the wild type (Figure 6.8). This plant, FT-96\_14 bolted seven days later and with 10.5 leaves more than the *AtCol0* wild type population (Table 6.4) so was chosen for sequencing of the *AtFT* gene to investigate whether a mutation was present.



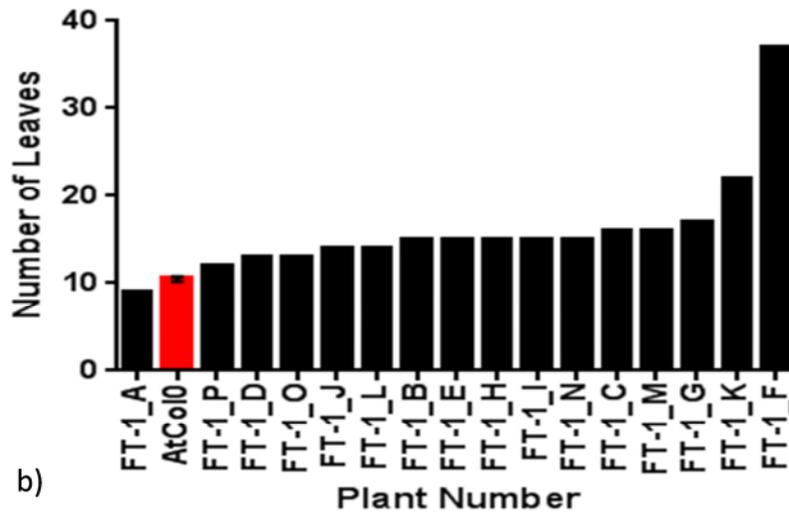
**Figure 6.8 – Flowering time of Arabidopsis plants transformed with CRISPR FT-96 construct.**

Number of leaves at bolting for Arabidopsis plants transformed with CRISPR FT-96 construct in experiment 1 (FT-96\_1-14). Black bars show transformed plants and red bar shows average flowering time for *AtCol0* wild type. Error bar represents standard error of *AtCol0* wild type mean flowering time.

Plants transformed with the CRISPR FT-1 construct flowered later than wild type in both experiments 1 and 2. In experiment 1, there were four lines flowering at least five days later with more than five leaves later than the wild type control and five lines flowering later with more than five leaves later compared to wild type in experiment 2, although all lines flower later in number of days than the wild type (Figure 6.9a and 6.9b, Table 6.4). Two lines in experiment 1 were noticeably later flowering than the wild type control by 24.5 leaves, although the number of days later is variable (Figure 6.10). This suggests that the CRISPR FT-1 construct has in many cases caused a mutation which has led to a disruption of FT function and delayed flowering. The late bolting plants from both experiments were investigated for possible mutations in the *AtFT* gene.



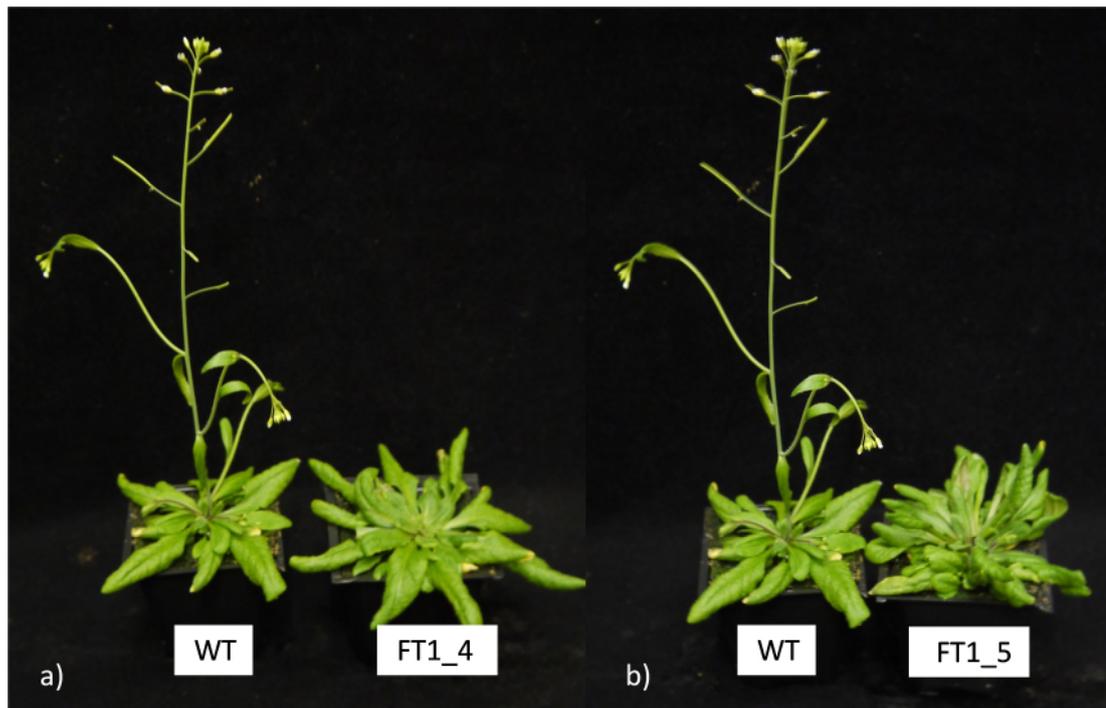
a)



b)

**Figure 6.9 – Flowering time of Arabidopsis plants transformed with CRISPR FT-1 construct in experiments 1 and 2.**

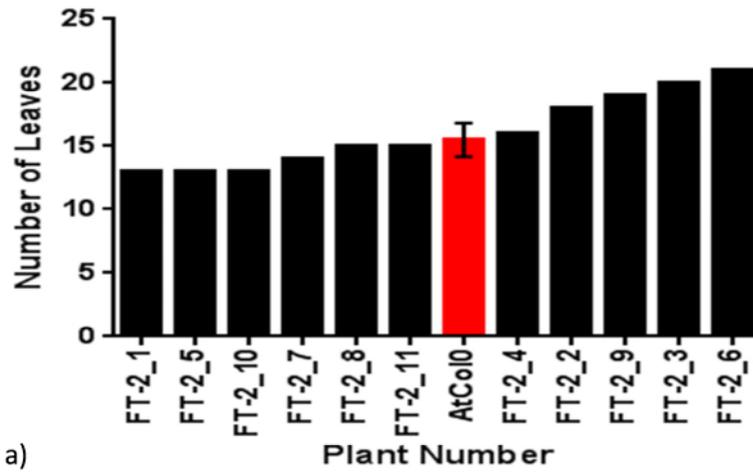
a) Number of leaves at bolting of Arabidopsis plants transformed with CRISPR FT-1 in experiment 1 (FT-1\_1-\_13). b) Number of leaves at bolting of Arabidopsis plants transformed with CRISPR FT-1 in experiment 2 (FT-1\_A-\_P). Black bars show transformed plants and red bar shows average bolting time for *AtCol0* wild type (n=4). Error bars represent standard error of *AtCol0* wild type mean flowering time.



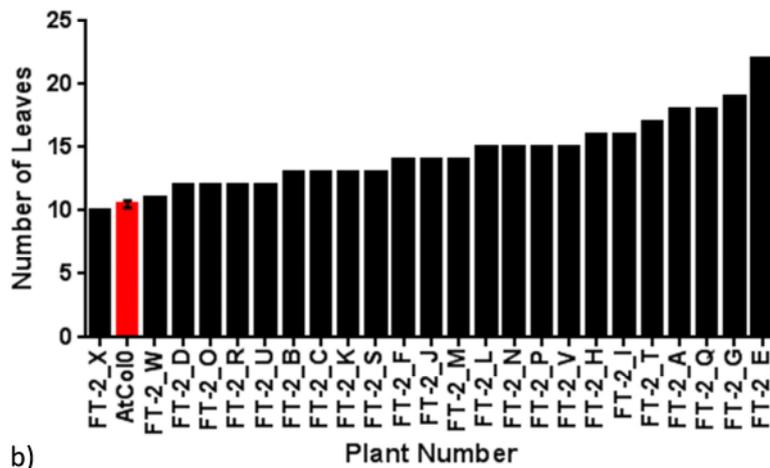
**Figure 6.10 – Flowering phenotype in LD of late bolting Arabidopsis lines FT1\_4 and FT1\_5 compared to wild type (WT).**

a) Late bolting phenotype observed for FT1\_4 compared to WT (*AtCol0*), b) Late bolting phenotype observed for FT1\_5 compared to WT (*AtCol0*). Photos taken 44 days after sowing.

The CRISPR FT-2 construct transformed Arabidopsis plants in experiment 1 showed that five transformed lines had more leaves than the *AtCol0* wild type average. Only one plant flowered with more than five leaves than *AtCol0*, and no plants flowered later in number of days than the wild type (Figure 6.11a, Table 6.4). Figure 6.12 shows plant FT2\_3 compared to WT where the late bolting phenotype is observed although it is only by 4.5 leaves. The transgenic lines from experiment 2 seem to be more affected by the CRISPR FT-2 construct with all but one plant showing a later phenotype by number of days and leaves, and seven plants where the difference compared to wild type is more than five leaves (Figure 6.11b).



a)



b)

**Figure 6.11 – Flowering time of Arabidopsis plants transformed with CRISPR FT-2 construct in experiments 1 and 2.**

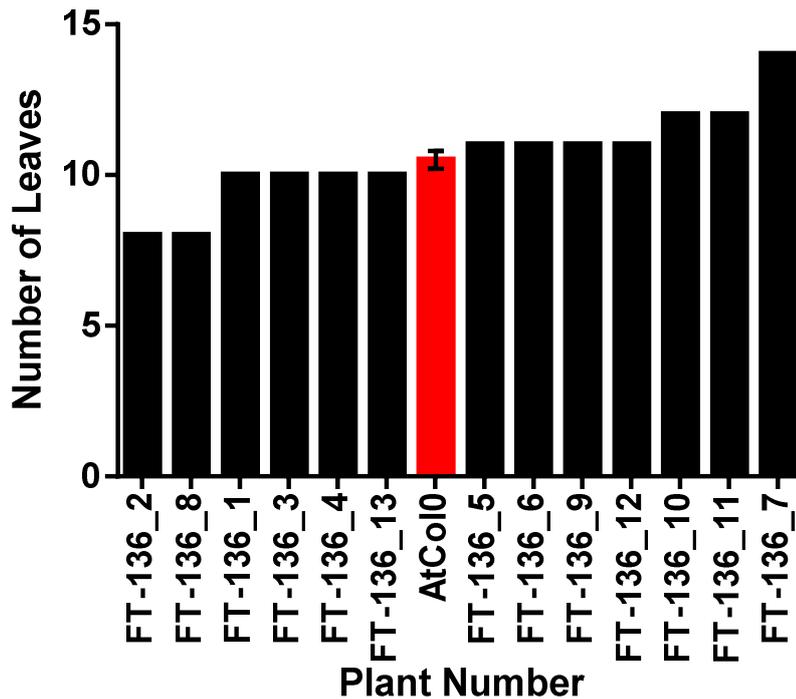
a) Number of leaves at bolting of Arabidopsis plants transformed with CRISPR FT-2 in experiment 1 (FT-2\_1-11). b) Number of leaves at bolting of Arabidopsis plants transformed with CRISPR FT-2 in experiment 2 (FT-2\_A-W). Black bars show transformed plants and red bar shows average bolting time for *AtCol0* wild type (n=4). Error bars represent standard error of *AtCol0* wild type mean flowering time.



**Figure 6.12 – Flowering phenotypes of Arabidopsis transformed line FT2\_3 compared to *AtCol0* wild type (WT).**

Late Flowering phenotype observed for FT2\_3 compared to WT (*AtCol0*). Photos taken 44 days after sowing.

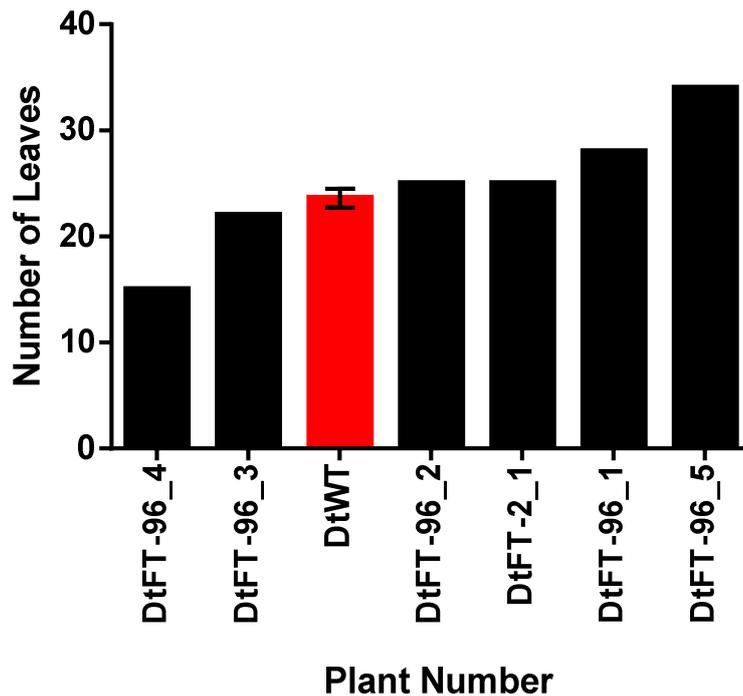
The flowering time results for plants transformed with CRISPR FT-136 construct show that six plants flowered with more leaves than the wild type with the latest of these being five days later with 3.5 more leaves (Figure 6.13). This suggests that the construct may not have been effective at introducing a mutation into the *FT* gene causing a major change in the expression of *FT*.



**Figure 6.13 – Flowering time of Arabidopsis plants transformed with CRISPR FT-136 construct.**  
 Number of leaves at flowering for Arabidopsis plants transformed with CRISPR FT-136 construct in experiment 2 (FT-136\_1-13). Black bars show transformed plants and red bar shows average flowering time for *AtCol0* wild type (n=4). Error bar represents standard error of *AtCol0* wild type mean flowering time.

#### 6.3.1.4. Rocket transformants

The flowering phenotype for Dt:CRISPR transformed plants was measured by leaf number at initiation of bolting. Dt:CRISPR FT-96 and Dt:CRISPR FT-2 transformed plants only successfully made it through the hygromycin selection in experiment 1, so there is no data available for experiment 2. However, due to plate contamination in experiment 1, no wild type plants from non-selection media plates were recovered, so the bolting data is a little unreliable as the transgenic lines are being compared to wild type plants grown on soil from seed. Figure 6.14 shows the bolting phenotype for the transformed plants where four plants are later than the wild type with one being more than five leaves later.



**Figure 6.14 – Bolting phenotype of *D. tenuifolia* plants transformed with CRISPR FT96 and CRISPR FT2 constructs.**

Number of leaves at initiation of bolting for rocket transformed with CRISPR FT-96 (FT-96\_1-5) or CRISPR FT-2 (FT-2\_1) in experiment 1. Black bars show transformed plants and red bar shows average bolting time for *Dt* wild type (n=3). Error bar represents standard error of *Dt* wild type mean flowering time.

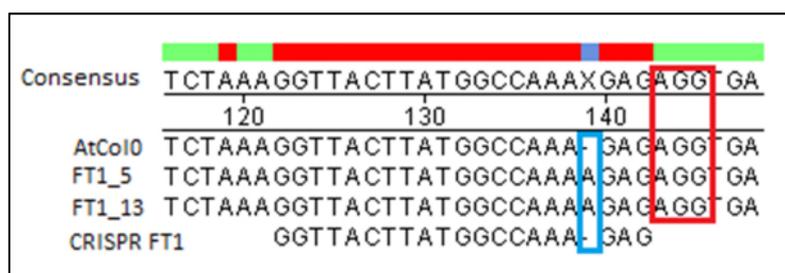
### 6.3.1.5. Sequencing results for *At*:CRISPR transformed plants

Leaf material from late flowering *At*:CRISPR transformed plants was collected and DNA extracted. PCR was used to amplify the full length *AtFT* gene (Appendix) and the products purified using a Qiagen PCR purification kit. Direct sequencing using primers covering the *AtFT* gene was carried out. Sequencing results were analysed using DNA Lasergene 9 core suite SeqMan and MegAlign packages to create and align contigs. The site in the *AtFT* gene where a mutation caused by the CRISPR/Cas construct is expected to be within the 20 nucleotide CRISPR target sequence. Therefore these regions of the FT sequence in the CRISPR transformants were aligned with the relevant CRISPR target sequence along with that part of the *AtFT* sequence from *AtCol0*. Table 6.4 shows all the potential CRISPR transformed lines which have the late bolting phenotype, along with the result of the direct sequencing.

**Table 6.4 – Summary of direct sequencing of FT PCR products from late bolting CRISPR transformed Arabidopsis plants**

Plant Name	Flowering time (No. of leaves later than WT)	Mutation detected?
FT-96_14	10.5	No
FT-1_3	9.5	No
FT-1_13	20.5	Single A Insertion
FT-1_4	24.5	Unclear
FT-1_5	24.5	Single A Insertion
FT-1_C	5.5	No
FT-1_M	5.5	No
FT-1_G	6.5	No
FT-1_K	11.5	Unclear
FT-1_F	16.5	Unclear
FT-2_6	5.5	No
FT-2_H	5.5	No
FT-2_I	5.5	No
FT-2_T	6.5	No
FT-2_A	7.5	No
FT-2_Q	7.5	No
FT-2_G	8.5	No
FT-2_E	11.5	No

The direct sequencing results revealed that despite the late bolting phenotype of 17 individual transformed plants, only two have a change in the *FT* gene introduced by the CRISPR/Cas9 construct. Both lines have an insertion of an A nucleotide, four nucleotides in from the PAM site, within the CRISPR FT-1 sequence (Figure 6.15). No other mutations were found in the rest of the *AtFT* gene in either the two CRISPR/Cas9 mutated lines or the 17 late lines.



**Figure 6.15 – Alignment of CRISPR FT-1 transformed Arabidopsis *FT* gene sequences showing mutations. Blue box shows the Adenine insertion, red box shows CRISPR PAM site.**

As some of the direct sequencing results did not detect a change in the *AtFT* gene, the CRISPR target region from seven late flowering FT-1 plants and four late flowering FT-2 plants were selected for cloning and sequencing. Four or five colonies from each cloned late flowering line were sequenced. The results revealed that there were no InDels present in the clones carrying products isolated from CRISPR FT-2 transformed plants FT2\_A, FT2\_E, FT2\_G or FT2\_Q. This suggests that the late flowering phenotype observed is not due to a mutation in the *AtFT* gene caused by the CRISPR FT-2 construct. It could instead be due to off-target mutations or by stress of antibiotic selection and transplanting to soil. Table 6.5 sets out the sequencing results found in clones from *AtFT* gene PCR products of CRISPR FT-1 transformed plants. The sequencing results for these were more encouraging. All late flowering CRISPR FT-1 transformed lines had at least one clone where a mutation (usually an A insertion or deletion) was present (Figure 6.16), and that the type of mutation was the same in all clones from the same line, except for FT1\_4. In this case, clones 1 and 3 showed no change in the sequence, clone 4 had an A nucleotide deletion, and clone 5 had a 22 nucleotide insertion (Table 6.5, Figure 6.16).

	90	100	110	120	130
Consensus	AAGGTTACTTATGGCCAA-----AAGAGAGGTGA				
9 Sequences					
CRISPR_FT_1_F	---GTTACTTATGGCCAA-----A-GAG				
AtCol0_3	AAGGTTACTTATGGCCAA-----A-GAGAGGTGA				
FT1_3_1	AAGGTTACTTATGGCCAA-----AAGAGAGGTGA				
FT1_13_4	AAGGTTACTTATGGCCAA-----AAGAGAGGTGA				
FT1_5_1	AAGGTTACTTATGGCCAA-----AAGAGAGGTGA				
FT1_G_2	AAGGTTACTTATGGCCAA-----GAGAGGTGA				
FT1_K_3	AAGGTTACTTATGGCCAA-----AAGAGAGGTGA				
FT1_F_2	AAGGTTACTTATGGCCAA-----AAGAGAGGTGA				
FT1_4_5	AAGGTTACTTATGGCCAA TGGCTTGGAGATTGGTGGCTTGGAGAGGTGA				

**Figure 6.16 – Alignment of CRISPR FT1 transformed and cloned Arabidopsis *FT* gene sequences showing mutations.**

Clone number is indicated by the last number in the sequence names. Blue box contains the Adenine insertion/deletion, green box contains the 22 nucleotide insertion. Red box contains CRISPR PAM site.

**Table 6.5 – Sequencing results from cloning PCR products from late flowering CRISPR FT-1 transformed plants.**

PCR products were cloned into pGEM-T easy and a number of clones sequenced using M13 primers. Sequencing results were aligned to *AtFT* and CRISPR FT-1 to see if any InDels were present.

<b>Plant Name</b>	<b>Colony Number</b>	<b>InDel Present?</b>	<b>Sequence Change</b>
FT1_13	1	Yes	Single A Insertion
	2	Poor quality sequence	
	3	Yes	Single A Insertion
	4	Yes	Single A Insertion
	5	Yes	Single A Insertion
FT1_F	1	No	
	2	Yes	Single A Insertion
	3	Yes	Single A Insertion
	4	Yes	Single A Insertion
	5	Yes	Single A Insertion
FT1_G	1	Poor quality sequence	
	2	Yes	Single A Deletion
	3	Yes	Single A Deletion
	4	No	
	5	No	
FT1_K	1	No	
	2	No	
	3	Yes	Single A Insertion
	4	No	
	5	No	
FT1_5	1	Yes	Single A Insertion
	2	Yes	Single A Insertion
	3	Yes	Single A Insertion
	4	Yes	Single A Insertion
FT1_3	1	Yes	Single A Insertion
	2	No	
	3	Poor quality sequence	
	4	No	
	5	Poor quality sequence	
FT1_4	1	No	
	2	Poor quality	

		sequence	
	3	No	
	4	Yes	Single A Deletion
	5	Yes	23 Nucleotide Insertion

### 6.3.2. Transient transformation

#### 6.3.2.1. Protoplast isolation

Protoplast isolation was done using the two methods outlined above, on both *Arabidopsis* and rocket leaf tissue. Table 6.6 is a summary of all experiments carried out over a five month period. Using method 1 (Yoo *et al.*, 2007), there is little difference in the number of protoplasts isolated from *Arabidopsis* compared to rocket. This showed that it was possible to isolate rocket protoplasts as well as *Arabidopsis* and in similar quantities. Once it became obvious that rocket protoplasts were not being efficiently transformed (section 6.3.2.2) or regenerating (section 6.3.2.3), the use of FDA to record the number of viable protoplasts became important. The use of FDA allows the viability of the isolated protoplasts to be estimated. The FDA is taken up by living cells and fluoresces green when exposed to UV/GFP light setting. The fluorescing protoplasts can then be counted on a haemocytometer using a fluorescent microscope. Optimisation of method 1 began by adjusting the concentration of mannitol in the enzyme solution, as this affects the osmolality of the protoplasts. The standard method used 0.4M mannitol, which works well for *Arabidopsis*. Concentrations of 0.4M, 0.6M, 0.65M and 0.7M were tried on rocket leaf material. The number of protoplasts isolated at each concentration was similar, but the viability varied. It is difficult to draw conclusions based on these results for the best concentration of mannitol to use as over several experiments at 0.6M mannitol, viability went from 0 to 20% and at 0.4M there was variation between 3.8%-30% and 0.65M and 0.7M concentrations were only tested once. The second method (Chatterjee *et al.*, 1988; Sikdar *et al.*, 1987; Sikdar *et al.*, 1990) was tried on rocket leaf material to see if the viability of cells produced would be better than method 1. Two attempts both yielded cells at a viability of  $\geq 30\%$  which was greater than seen with method 1. However, in subsequent isolations, the viability was only 3.5%. This can probably be accounted for by the change of

pectinase enzyme used, as the activity of pectinase (and the other enzymes in the procedure) can vary between packages and manufacturers.

**Table 6.6 – Summary of protoplast isolations from Arabidopsis and rocket leaf material.**

Standard method 1 contains 0.4M mannitol, brackets in table show what the mannitol concentration was changed to in those isolations (n/d – viability not determined) (Appendix D.1)

Date of Isolation	Arabidopsis		Rocket		
	Method	Number of Cells	Method	Number of Cells	Viability
1/5/14	1	300,000 cells/ml	1	164,000 cells/ml	n/d
7/5/14	1	242,000 cells/ml	1	64,000 cells/ml	n/d
13/5/14	n/a	n/a	1	258,000 cells/ml	n/d
19/5/14a	1	210,000 cells/ml	1	142,000 cells/ml	n/d
19/5/14b	1	204,000 cells/ml	1	154,000 cells/ml	n/d
21/5/14	n/a	n/a	1	266,000 cells/ml	n/d
27/5/14	1	238,000 cells/ml	1	264,000 cells/ml	n/d
14/7/14	1	228,000 cells/ml	1	270,000 cells/ml	n/d
28/7/14			1 (0.6M)	224,000 cells/ml	n/d
11/8/14			1 (0.65M)	230,000 cells/ml	12.35%
11/8/14			1 (0.7M)	230,000 cells/ml	15.14%
13/8/14			1 (0.6M)	388,000 cells/ml	20%
22/8/14			1	Not measured	30%
26/8/14			1 (0.6M)	242,000 cells/ml	0
26/8/14			2	462,000 cells/ml	37%
9/9/14			1	350,000 cells/ml	21.37%
11/9/14			2	262,000 cells/ml	30%
15/9/14			1	228,000 cells/ml	11.5%
15/9/14			1 (0.6M)	201,200 cells/ml	15%

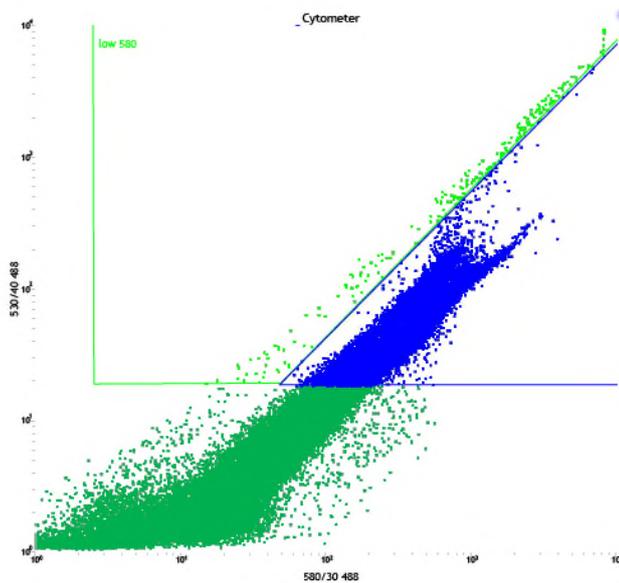
17/9/14			2	266,680 cells/ml	3.5% (new pectinase)
22/9/14			1	275,000 cells/ml	3.8%
25/9/14			2	246,680 cells/ml	3.5% (new pectinase)

### 6.3.2.2. Protoplast transformation

Arabidopsis protoplasts isolated using method 1 (Yoo *et al.*, 2007) were used in single PEG-mediated uptake transformations with GFP and double transformations with GFP and mCherry or CRISPR plasmids. This was done by gently mixing 100  $\mu$ l protoplasts (200,000 cells/ml for single transformation and 400,000 cells/ml for double transformation) with a PEG-calcium solution and plasmid DNA. The solution was incubated at room temperature for 30 min before stopping the reaction with cold W5. Following centrifugation, the supernatant was removed and the protoplasts resuspended in W1 solution, covered with Nescofilm and left at room temperature for 24 hours. An aliquot of protoplasts was taken the following day and observed under a fluorescent microscope. Transformation of Arabidopsis protoplasts was attempted twice with a single plasmid and four times with a double transformation (appendix). It was found that the single transformations worked as GFP fluorescence was observed (Figure 6.17). Three out of the four double transformations attempted were found to work as fluorescence was observed in some cells. This showed that the GFP plasmid had successfully been transformed into the cells, but there was no way of testing whether the CRISPR plasmid was also transformed into the cells. Both results meant that when Arabidopsis protoplasts were put through the FACS machine, the GFP fluorescing ones could be detected and sorted (Figure 6.18). Therefore it was shown that Arabidopsis protoplasts could be transformed with CRISPR constructs as single transformations, or in double transformations with CRISPR constructs together with a GFP plasmid.



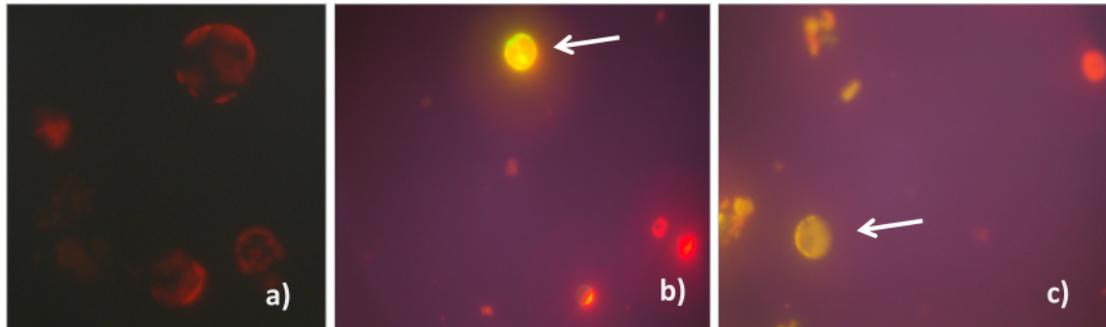
**Figure 6.17 – Transformed Arabidopsis protoplast fluorescing with GFP**  
(isolated cells from 19/5/14)



**Figure 6.18 – FACS result for Arabidopsis protoplasts transformed with GFP (isolated cells from 27/5/14).**  
X axis shows fluorescence emitted at 580nm +/- 15 from 565 to 595. When excited with the 488nm laser it represents autofluorescence. Y axis shows fluorescence emitted at 530nm +/- 20 from 510 to 550. When excited with the 488nm laser it represents GFP fluorescence. Dark green – small cells or debris, blue – good size cells, not transformed, light green – good cells, GFP transformed

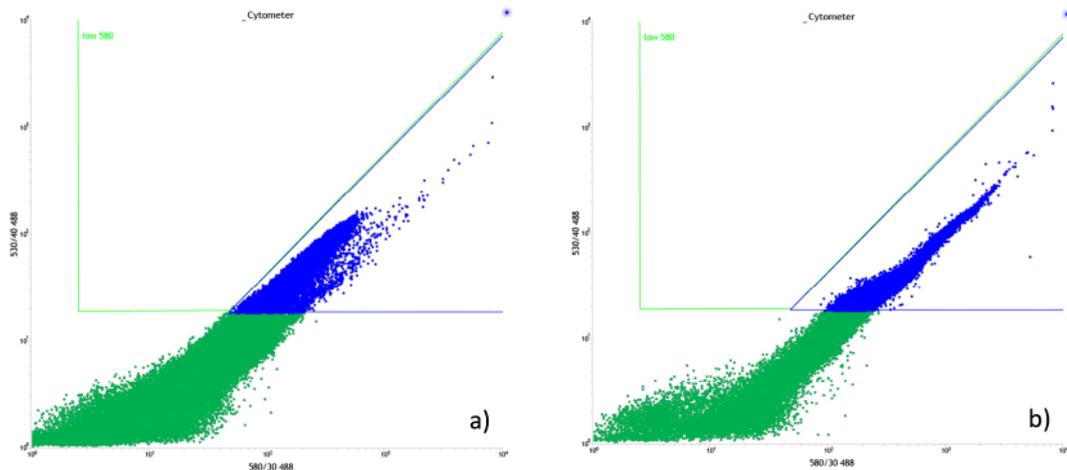
Attempts at single or double transformations of rocket protoplasts achieved little success. Figure 6.19a shows transformed cell lines where no fluorescence is seen. In one experiment (21/5/14), cells transformed with GFP and mCherry (Figure 6.19b), or YFP and mCherry (Figure 6.19c) show the expected fluorescence, suggesting that the technique works, but this was not able to be replicated at a later date. Unfortunately, transformations using a GFP and CRISPR construct did not yield any fluorescing protoplasts. GFP-transformed rocket cells were analysed using the

FACS machine (Figure 6.20). The results show that many cells were detected by the machine, but that the results for the non-transformed and transformed cells were the same, that no GFP fluorescence was detected and so the cells could not be sorted.



**Figure 6.19 – Transformed rocket protoplasts.**

a) Rocket protoplasts transformed with GFP, but no green fluorescence observed (isolated cells from 19/5/15). The red cell colour is due to autofluorescence of the chloroplasts. b) Rocket protoplasts transformed with GFP and mCherry. White arrow points to a cell which has been transformed with both plasmids as it is fluorescing both green and red (isolated cells from 21/5/15). c) Rocket protoplasts transformed with YFP and mCherry. White arrow points to a cell which has been transformed with both plasmids as it is fluorescing both yellow and red (isolated cells from 21/5/15)



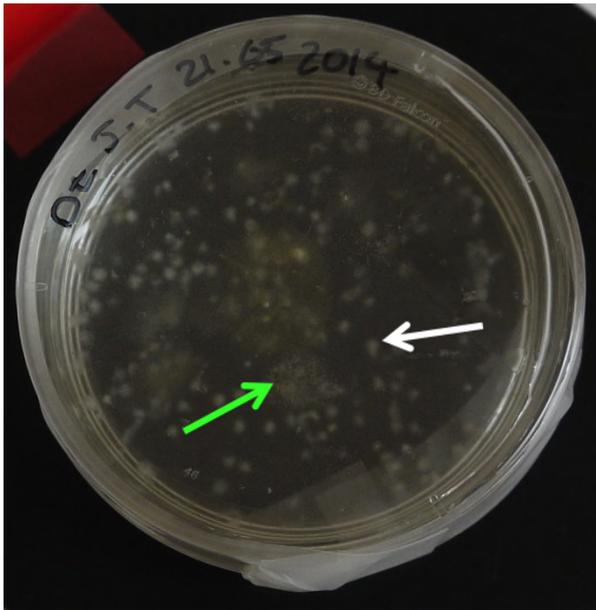
**Figure 6.20 – FACS result for transformed and non-transformed rocket protoplasts from protoplast isolation 27/05/14.**

a) Non-transformed rocket protoplast control sample, no GFP transformed cells present. b) Rocket protoplasts transformed with GFP, no GFP transformed cells present. X axis shows fluorescence emitted at 580nm +/- 15 from 565 to 595. When excited with the 488nm laser it represents autofluorescence. Y axis shows fluorescence emitted at 530nm +/- 20 from 510 to 550. When excited with the 488nm laser it represents GFP fluorescence. Dark green – small cells or debris, blue – good size cells, not transformed, light green – good cells, GFP transformed

### 6.3.2.3. Regeneration of protoplasts

Both transformed and untransformed Arabidopsis and rocket protoplasts, were plated a number of times over the course of the experiment onto media M8PS2 to try to

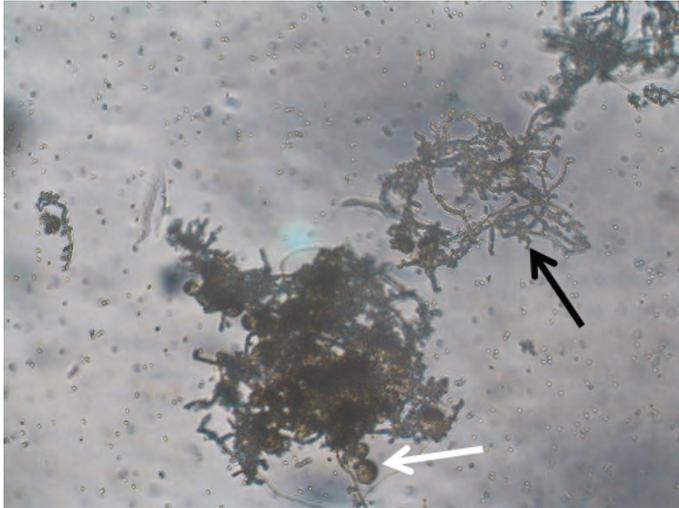
regenerate cells into whole plants. The number of cells plated varied from 10,000-400,000 cells/ml to see if the quantity of protoplasts used had any effect of the number that were able to regenerate, or if clumping of cells due to the high number would inhibit regeneration. After 10 days incubation at 24°C in the dark, a few plates contained small cell masses suggesting that proliferation had occurred, so these were transferred onto media M8PS21 (Figure 6.21).



**Figure 6.21 – Rocket protoplasts at regeneration day 10.**

White arrow points to a clump of cells which could be protoplasts that have divided to 16-64 cell callus, so were transferred to media M8PS21. Green arrow points to possible fungal contaminants.

The progress of one plate containing *Arabidopsis* cells transformed with GFP at day 16 was observed under a microscope (Figure 6.22). This showed some round cells which were possibly the protoplasts still present, although undivided and a large amount of filamentous material which could have been fungal contamination. Many plates were discarded by day 20 due to the presence of a greenish filamentous type mass of material. According to Sikdar et al. (1990), by day 20 a globular or heart-shaped embryo should be present. This was never observed during these experiments due to fungal contamination.



**Figure 6.22 – Arabidopsis protoplasts at regeneration day 16.**

White arrow points to a clump of protoplast cells. Black arrow points to possible fungal contamination.

## **6.4. Discussion**

### **6.4.1. Production of CRISPR transformed plants**

*A. tumefaciens* and floral dipping was used to transform the CRISPR/Cas9 construct into *Arabidopsis* and rocket. Four different CRISPRs designed to *FT* were transformed in and all produced T<sub>1</sub> plants after hygromycin selection. Late flowering lines were recovered which is what was expected if the CRISPR/Cas9 had functioned to cut the *FT* gene. Upon analysis of the lines by direct sequencing, only two showed a sequence change of an additional Adenine inserted at the expected position three nucleotides upstream from the PAM (Table 6.4, Figure 6.15). However, after cloning the PCR amplified fragments from FT-1 and FT-2 late bolting plants, more were found to have mutations within the CRISPR target sequence in the *AtFT* gene (

Table 6.5, Figure 6.16). This was only true however for FT-1. The cloning revealed that all the late flowering plants tested here contained a mutation in at least one of the sequenced clones. For two lines, FT-1\_13 and FT-1\_5, all clones obtained which provided good sequence contained a single Adenine insertion, suggesting that the mutation could be homozygous in the plant. This would be due to the CRISPR/Cas9 causing the DSB in the germ cell from which the plant was derived. More clones would need to be sequenced to verify this and the segregation of the T<sub>2</sub> generation analysed. Not all clones obtained from lines FT-1\_F, FT-1\_K and FT-1\_3 contained the additional Adenine. This suggests that the insertion is heterozygous in these plants and analysis of T<sub>2</sub> segregation would verify this. The Adenine insertion causes a disruption in the open reading frame in the *FT* gene, leading to a stop codon being introduced five codons downstream from the insertion site. Therefore after transcription, the translation would lead to a 39 amino acid peptide of the *AtFT* gene being produced and the plant would no longer have a functional *FT* gene. This could explain the late flowering phenotype observed in these plants. In two lines, FT-1\_G and FT-1\_4, the mutation seen in some clones is a deletion of an Adenine. The sequencing of the other clones from those lines gave WT sequence, so the plants are heterozygous for the Adenine deletion. The Adenine deletion in the *FT* gene disrupts the open reading frame, and a stop codon is introduced two codons down from the deletion site. This would mean that the plant would no longer have a functional FT protein, explaining the late bolting phenotype observed in these plants. One sequenced clone for FT-1\_4 also has a large insertion of 22 nucleotides where the same Adenine as in the Adenine deletions is also deleted. The plant is likely to be heterozygous for this insertion as it was only found in one clone. Due to the deletion of the Adenine and the insertion of a random 22 nucleotides, the open reading frame is maintained, so the *FT* gene would continue to be translated. The insertion would introduce seven new amino acids (Arginine is replaced by a tryptophan and is followed by leucine, glycine, aspartic acid, tryptophan, tryptophan, leucine and glycine before re-joining the correct amino acid sequence) (Figure 6.23). This could cause a change in the protein structure, perhaps affecting its binding affinity and reducing its functionality, so making the plant late bolting.

Sequence Nam	< Pos = 11
- ■■■■■ +	
<input checked="" type="checkbox"/> Consensus 2 Sequences	VSRVVGDVLDPFNRSI TLKVITYGQXXXXXXXXXXEVTNGLDLRPSQVQNKPRVEI GGE
At FT	VSRVVGDVLDPFNRSI TLKVITYGQR- - - - - EVTNGLDLRPSQVQNKPRVEI GGE
FT1_4_5	VSRVVGDVLDPFNRSI TLKVITYGQWLGDWMLGEVTNGLDLRPSQVQNKPRVEI GGE

**Figure 6.23 – Alignment of AtFT protein sequence against the translated FT1\_4\_5 clone sequence showing the additional amino acids.**

The FT1\_4\_5 clone sequence of the *FT* gene was translated in frame from the same start codon as found in *AtFT*. This was then aligned to the *AtFT* amino acid sequence to show where the additional sequence due to the CRISPR/cas9 FT-1 construct has changed the translation. Red line shows identical sequence, blue line shows where there is a difference.

The late flowering phenotype observed in the other transformed plants may be due to off target site mutations caused by the CRISPR cutting elsewhere in the genome. The 12 nucleotides upstream from the PAM are the most crucial to the specific targeting of the CRISPRs (Pattanayak *et al.*, 2013), so these were blasted against the *A. thaliana* genome for all the CRISPR target sequences that were used in these experiments to see what may be possible off-target sites (Table 6.7).

**Table 6.7 – Blast results for 12 nucleotide sequence from PAM site for all CRISPRs to investigate possible off-target sites.**

Blastn used to search for 100% identity matches to CRISPR target sequences in *Arabidopsis thaliana* genome. Red text denotes where sequence has been found in blasting full CRISPR sequence as well as the 12 nucleotides blasted here.

CRISPR sequence	Blastn results (100% homology)
FT-136 - TCTCAGGTCAA	UB-like protease 1B
	leucine-rich receptor-like protein kinase
	pentatricopeptide repeat-containing protein
	aminopeptidase
	sucrose-phosphatase 2
	Afadin/ $\alpha$ -actinin-binding protein (73% homology to 20 bp FT-136 sequence)
	adenine nucleotide $\alpha$ -hydrolase domain-containing protein kinase
	actin-related protein 8A and actin-related protein 8B (ARP8) gene
	G-type lectin S-receptor-like serine/threonine-protein kinase
	ARM repeat superfamily protein
	ethylene-responsive transcription factor RAP2-4
	RST1 (RST1)
FT-1 - ATGGCCAAAGAG	phosphatidate cytidyltransferase
	nodule inception protein-like protein 8

	AT4G19120
	long chain acyl-CoA synthetase 4
	uncharacterized protein
	S-adenosyl-L-methionine-dependent methyltransferase
	putative methyltransferase PMT21
	phosphoenolpyruvate carboxylase family protein
	S-adenosyl-L-methionine:jasmonic acid carboxyl methyltransferase
	jasmonic acid carboxyl methyltransferase
	long chain acyl-CoA synthetase 4 (LACS4)
	cytochrome P450, family 87, subfamily A, polypeptide 2
	putative serine/threonine protein kinase
	monomeric G protein SGP2
	mRNA for disease resistance like protein
	TIR-NBS-LRR class disease resistance protein
	non-specific phospholipase C2
FT-2 - AGCAACCCTCAC	<b>TWIN SISTER of FT (68% homology to 20 bp FT-2 sequence)</b>
	glyoxal oxidase-related protein
	clone U11796 putative flowering signals mediating protein FT (At1g65480)
	At3g28750 unknown protein
FT-96 - AAAGGTTACTTA	glutamine synthetase 2
	3-epi-6-deoxocathasterone 23-monooxygenase
	pentatricopeptide repeat-containing protein
	At4g36380
	<b>cytochrome P450 (68% homology to 20 bp FT-2 sequence)</b>
	Subtilase 5.4
	protein IQ-domain 16
	putative glutamate-ammonia ligase precursor, chloroplast (At5g35630)
	DM2A, DM2B, DM2C, DM2D, DM2E, DM2F, DM2G, and DM2H genes
	protein TRANSPARENT TESTA 12
	nucleotide-sugar transporter family protein

From the Blast results, all four CRISPRs have many possible off-target sites, which may be why some lines were late flowering but did not have a mutation in the *FT* gene. The most interesting of these is FT-2. Late bolting plants were seen for this CRISPR transformed line, but none had a mutation in the CRISPR target site of the

*AtFT* gene. When the off-target sites were checked, *AtTSF* (*TWIN SISTER OF FT*) was found to be one of the hits. When the CRISPR FT-2 whole 20 bp sequence was blasted, there was only a 68% homology to the *AtTSF* sequence, where just 13 nucleotides matched out of the 20, which is why this sequence was still considered a viable option. *AtTSF* is also found in the flowering time pathway and acts in the same way as *FT* but with a weaker effect. A mutation could therefore be present in the late bolting FT-2 plants in the *AtTSF* gene instead of *AtFT*. This should be verified by sequencing and cloning the *AtTSF* gene as performed for *AtFT* to see if this is the case. The Arabidopsis clone U11796 putative flowering signals mediating protein FT was also found to match 100% of the 12 nucleotide sequence of CRISPR FT-2 but wasn't visible when BLASTed using the 20 bp sequence. It still could be a credible off-target site where a mutation introduced could alter the flowering phenotype. This could be verified by cloning and sequencing the gene from the late flowering FT-2 plants. For the FT-96 and FT-136 CRISPRs there are no immediately obvious flowering related off-target sites. The late flowering in those plants with no sequence change in the *FT* gene could also be the result of stress on the plants having been through the hygromycin selection and transferred to soil, but this would become evident when they are re-tested more thoroughly in the T<sub>2</sub> generation. There was a difference in flowering time between the *AtCol0* WT plants in experiments 1 and 2 by five leaves, so this may have occurred in the transformed plants between experiments as well. The number of plants which were wild type bolting which came through the hygromycin screen could be due to the plant being transformed by the resistance gene containing vector, but where the CRISPR failed to cut the *FT* gene, or a cut was made but the DNA repair restored the DSB correctly, so no mutation occurred, or introduced a mutation in an off-target site that didn't affect flowering.

Further analysis of the mutated lines is necessary to test whether the mutations are carried through to the T<sub>2</sub> generation (Jiang *et al.*, 2014). This could also shed light on whether late flowering lines found not to have a mutation in the *FT* gene produce late flowering progeny. The T<sub>2</sub> generation of the late flowering lines where the cloning of the *AtFT* region showed different results could make the mutation and phenotype much clearer through segregation of these traits. Putative late bolting lines were also seen in rocket, but due to the lack of proper WT controls which had been through the media selection stage, this data is not very well supported. It was

also not possible to test whether there was a mutation present at all in the CRISPR site as the isolation of *DtFT* from gDNA isn't feasible at this stage. Further work is needed to isolate a full length *DtFT* gene from gDNA and therefore design working primers to this to amplify the region where the CRISPRs are targeted to. If there is an introduced mutation in the *DtFT* in the late bolting lines this would show that floral dipping is a viable method of transformation in rocket, despite the drawbacks of low seed production, and that the CRISPR/Cas9 system can also function in rocket. Both would give huge advantages to the continuation of breeding programmes in rocket. The work in this study using Arabidopsis has shown that different mutations types have been introduced into the *AtFT* gene using the CRISPR/Cas9 system. A deletion of an Adenine produced the desired late flowering phenotype, which if replicated in rocket would qualify it as a safe GMO for sale in the USA and possibly Europe if the legislation changes.

#### **6.4.2. Transient expression in protoplasts**

The experiments performed to isolate protoplasts from Arabidopsis and rocket using two different methods have shown that protoplasts can be made from leaf material of each species in similar quantities. However, it was found that using FDA as a measure to estimate the viability of the isolated protoplasts was vital if transformation and regeneration were to be successfully carried out afterwards. The viability was not measured for Arabidopsis, but transformation of the GFP plasmid was shown to be effective by microscopy and FACS, so enough live cells must have been present for this to work. Viability in rocket protoplasts varied quite a lot within and between methods and isolations from 0 to 37%, but the onward experiments showed little success. To make a recommendation for which method to use in future experiments, further testing and optimisation of each method 1 and method 2 is needed due to the low viability of protoplasts. Sikdar et al. (1987) were able to regenerate protoplasts into whole plants using protoplast solution where the viability was 80%. Protoplast viability of a maximum of 37% is therefore not high enough for subsequent transformation and regeneration experiments to work effectively.

Transformation of Arabidopsis protoplasts using the PEG calcium method outlined in section 6.2.3.2 with GFP, YFP or mCherry plasmids in either single or double transformations was shown to work with fluorescence microscopy and FACS. Where a double transformation was performed with GFP or YFP and CRISPR/Cas9,

the fluorescence of the GFP or YFP was observed, but until callus or plants are regenerated it is not possible to test whether the CRISPR/Cas9 had also been transformed in. In rocket, only in one experiment did the protoplast transformation using the PEG calcium method work as shown by fluorescence microscope observation. This could be due to the low viability of cells after the isolation and the efficiency of the PEG calcium transformation. If both of these were optimised, the transformation efficiency may improve and be repeatable. However, other methods such as electroporation of protoplasts could be tried (Velasquez *et al.*, 2012) to see if rocket is amenable to transformation in this way. The same problem of identifying whether the CRISPR/Cas9 plasmid had been transformed in is also true in rocket. The construct in pCAMBIA1300 has a hygromycin resistance gene within it, so a method could be optimised to treat the transformed protoplasts with hygromycin before plating for regeneration. Direct DNA extraction of transformed protoplasts could also be tried before PCR performed to see if the construct is present and if an InDel is present in the *FT* gene (Li *et al.*, 2013b). If the protoplasts were treated with hygromycin after transformation, this would increase the concentration of transformed cells in the extraction. If relying on a fluorescent tag for identification of transformed protoplasts is to be used effectively, then building a CRISPR/Cas9 construct with GFP incorporated in it would be sensible. Double transformations do not work as effectively as single transformations. So protoplasts transformed with both the CRISPR/Cas9 and GFP plasmids will produce a mixture of cells with one plasmid, or the other, or both. When enrichment of GFP carrying cells by FACS is used, the protoplasts collected at the end will be a mixture of GFP single transformants and GFP and CRISPR/Cas9 double transformants, and therefore loses the percentage of CRISPR/Cas9 single transformed cells. If the CRISPR/Cas9 construct has the GFP tag within it, a single transformation can be used, and all cells captured in FACS will contain the desired CRISPR/Cas9 plasmid.

The use of a FACS machine to detect and separate the GFP transformed protoplasts is an effective way of isolating GFP labelled cells from a solution for further analysis and DNA/RNA extraction. However, it is not an ideal method to use for these experiments as the protoplasts carrying the GFP-expressing plasmid were required for regeneration. To try and prevent the contamination of the regenerating cells, the cells must be kept in sterile conditions. The sheath fluid which pumps round the FACS machine, carrying the cells through the lasers for detection and separation,

isn't sterile and the machine isn't set up to run a sterilising solution through prior to the protoplast carrying sheath fluid. Therefore, contaminants could be picked up from this stage. Also, for the regeneration to be successful, the cells must be viable in order to be able to divide and proliferate. The mechanical stress on the protoplasts running through the machine, and the force at which they hit the collecting solution after separation, can cause the cells to rupture. This could be improved by increasing the number of transformed cells in the starting solution to compensate for those which are ruptured in the sorting and by optimising the collection fluid osmolality and height from which the sorted cells are dropped from.

The regeneration of *Arabidopsis* and rocket protoplasts was undertaken with the aim of producing plants which had been modified by the CRISPR/Cas9 system in the *FT* gene. The regeneration method for this outlined by Sikdar et al. (1990) was used as it had been optimised for *D. muralis*. Difficulties were found in replicating the success observed by this group and no cell lines made it past day 25. The procedure was carried out in a class II cabinet and sterile technique improved over time, but with no effect. The protoplasts tended to clump together when plated, so different concentrations were applied from  $10^4$ - $10^6$  to try and improve this. Plates were consistently contaminated, so all media and components were sterilised by autoclaving or filter-sterilisation. This should be effective enough to overcome contamination, but it didn't improve the problem. Milton® sterilising tablets (Troloxene sodium) were dissolved in water and the solution added to each media to provide further sterilisation. This did prevent fungal and bacterial contaminants from growing, but no regeneration of protoplasts was seen in these plates, so it may also have had a negative effect on the plant cells present. Overall, the regeneration of *Arabidopsis* and rocket protoplasts was unsuccessful, despite changing cell concentrations plated, improvements in sterile technique and ensuring all media was prepared correctly and sterile.

## Chapter 7. General Discussion

### 7.1. Isolation and characterisation of Arabidopsis flowering time gene homologues in rocket

The flowering time pathway in Arabidopsis is a complex network of genes which responds to environmental and endogenous signals to bring about the competence to flower. The application of this knowledge to crop plants has begun, with key genes and pathways being elucidated in monocot and dicot plants and both long day and short day responsive plants. Rocket is a relatively new crop plant to the UK and the understanding of its floral pathways could help to tackle one of the issues associated with its cultivation, namely early bolting. Five key flowering pathway genes were chosen for study in rocket as they lie at convergent points between flowering pathways in Arabidopsis. These were *CO*, *GI*, *FLC*, *SOC1* and *FT*. Each gene was isolated from rocket cDNA using PCR and RLM-RACE and cloned into *A.*

*tumefaciens* for complementation studies in relevant Arabidopsis mutants. The full length genes including introns were isolated via PCR using primers designed to 5' and 3' UTR sequence outside of the CDS and sequenced. Real time primers were designed to regions within the gene enabling the change in gene expression to be observed over time in both diurnal and developmental time course experiments. The experiments undertaken were successful in isolating the CDS for all the genes of interest including two *FT*-like genes, *DtFT* and *DtTSF*. All genes were transformed into their equivalent Arabidopsis mutants and *DtCO*, *DtSOC1*, *DtFLC* and *DtFT* were found to complement their respective mutation and thus be functional orthologues. The complementation using *DtGI* failed to yield any T<sub>1</sub> plants after the BASTA selection so the experiment will need repeating. *DtTSF* was unable to complement the *ft-10* mutant but the flowering time phenotype was altered. It produced very late flowering plants with aberrant flowers, unlike *DtFT* which produced the expected early flowering phenotype. In hindsight, it would have been useful to also transform isolated CDS from each gene in Arabidopsis into its equivalent mutant plant as a direct comparison for the rocket complementation results. To gain further understanding of the rocket flowering time gene homologues, the expression of each gene was investigated over diurnal and developmental time courses. It was found that *DtCO*, *DtGI*, *DtSOC1*, *DtFT* and

*DtTSF* were diurnally regulated whilst *DtFLC* was not. This is consistent with their Arabidopsis homologues (Fowler *et al.*, 1999; Kobayashi *et al.*, 1999; Srikanth and Schmid, 2011; Suarez-Lopez *et al.*, 2001), providing more evidence that each is a true orthologue (except *DtTSF*). The developmental time course was also informative. *DtCO*, *DtFT* and *DtTSF* showed an increase in expression at the time point before bolting was initiated, whereas *DtFLC* showed a slight decrease in expression before bolting. The expression of *DtGI* and *DtSOC1* appeared not to change before bolting occurred. It would be worth repeating both time course experiments as only one biological replicate was used and results did not always show clear patterns, making interpretation difficult. The developmental time course was also reliant on all genes tested being expressed most highly at the point of material collection (ZT15.5), which according to the diurnal expression results was not the case. It should therefore be repeated again using a more optimal collection time for all genes. From the evidence gathered so far, it seems as though all the genes investigated have the same role in the flowering time pathway as their Arabidopsis homologues. This is except for *DtTSF* which is a repressor rather than an inducer of flowering according to the *ft-10* complementation results. This suggests that rocket has multiple *FT*-like genes with opposing functions, like the *FT* genes discovered in sugar beet (Pin *et al.*, 2010) and onion (Lee *et al.*, 2013). Future experiments could include diurnal time course experiments using short days to test if the expression patterns are still similar to Arabidopsis SD expression. A diversity set for *Diplotaxis* could be gathered and the number of days to bolt measured to see if there are differences within this population compared to the *Dt* WT used in this project. Lines bolting earlier and later than WT could then be used to investigate changes in expression over developmental and diurnal time courses of the six flowering time genes isolated, and each of these genes sequenced in the early and late bolting lines to see if any SNPs are present in these genes which may explain the difference in bolting phenotype. Other genes involved in the flowering pathway could be isolated and function investigated such as *FKF1*, *FRI*, *FD*, *miR156*, *miR172*, *SVP*, *TFL*, autonomous pathway genes and key circadian clock genes (for example *PRR* (Cavauiolo *et al.*, 2015)). The interactions between these and the *Dt* flowering time genes that we have already isolated could then be investigated.

## 7.2. Understanding the vernalization and photoperiodic requirements of rocket

The vernalization and photoperiodic pathways are two key pathways involved in the perception of and response to environmental conditions. Different responses to vernalization and photoperiod have led to different ecotypes with *Arabidopsis* (Strange *et al.*, 2011) and different varieties of crop plants such as wheat and oilseed rape (Bluemel *et al.*, 2015; Srikanth and Schmid, 2011). Understanding the vernalization and photoperiodic requirements of rocket could lead to breeding of new varieties which respond differently, and therefore have different phenotypes.

Studies done on the vernalization requirements of *Arabidopsis* and *Brassica* showed that within the *Brassicaceae* family there are differences in vernalization treatment length, temperature and in the growth stages that are sensitive to vernalization (Guo *et al.*, 2004; Jung and Mueller, 2009; Nordborg and Bergelson, 1999). Therefore experiments were set up with rocket to test two vernalization temperatures, at two different growth stages, over three treatment lengths. The results revealed that rocket does not have a vernalization requirement as the cold treatment did not reduce the time the plants took to bolt consistently, under either temperature. This could be due to its origins in the Mediterranean where the lack of a vernalization response would not be a disadvantage. This is similar to *B. napus* varieties grown in the Mediterranean where the lack of a vernalization requirement is important to its early flowering phenotype (Dahanayake and Galwey, 1999). However, when the expression of *DtFLC* was analysed, in both 4°C and 10°C seed vernalization experiments, after just one week of cold the level of *DtFLC* expression was reduced to very low levels and remained low for the duration of the treatment. This was also observed in the 4°C four week old plant vernalization experiment where *DtFLC* expression was reduced in the plant in the cold treatment and remained low even when the plants were returned to ambient temperatures. These results suggest that the vernalization pathway is still present in rocket as *DtFLC* expression was reduced, but there was no subsequent reduction in the bolting time of the treated plants, so this isn't enough to induce flowering in rocket. This may be due to some lesion in the flowering time pathway between *DtFLC* and the floral integrator genes. It could also be that *DtFLC* is not expressed highly enough to repress flowering in rocket as it flowers rapidly, so the reduction in expression during vernalization is not significant

enough to deviate from its normal flowering habit. This would be similar to Arabidopsis ecotypes *Ler*, *Da* and *Shakhdara* which have a low expression of *AtFLC* but flower rapidly and are classified as summer annuals. It was found that the *AtFLC* in *Ler*, *Da* and *Shakhdara* did not contain any mutations in the CDS compared to *Col0*, but the low expression was due to mutations in intron 1 (*Ler*) or 5' UTR (*Da* and *Shakhdara*) sequences (Michaels *et al.*, 2003). The work done in this project using the *DtFLC* CDS to complement the Arabidopsis *FRI+flc3* mutant showed that *DtFLC* is a functional orthologue of *AtFLC*. Further experiments to express an active *AtFLC* allele in rocket and observe the effect on bolting time could be done as carried out in Michaels *et al.* (2003). Spring type *B. napus* does not have a vernalization requirement (Rakow *et al.*, 2011) and this was found to be due to the sequence difference at the *FLC* locus on chromosome A10. Hou *et al.* (2012) analysed *BnFLC* expression in winter type Tapidor and semi-winter type Ningyou7 and found that *BnFLC* expression dropped over vernalization in both varieties, but much more quickly in the Ningyou7, which is less sensitive to vernalization than Tapidor. The difference in vernalization responsiveness between the two varieties tested, and other spring types was shown to be a miniature inverted-repeat transposable element (MITE) which was inserted upstream of the start codon of the *BnFLC.A10* gene in vernalization sensitive winter type varieties such as Tapidor. It would be interesting to insert this MITE sequence into an equivalent location upstream of *DtFLC* to see if there is a change in the vernalization requirement of rocket. It is theorised that the non-vernalizing spring types of *B. napus* evolved into the winter types when the crop was moved from its origins around the Mediterranean to northern latitudes due to the insertion of the MITE near the *BnFLC.A10* locus through selective breeding (Hou *et al.*, 2012).

The vernalization experiments carried out here were pilot experiments to investigate if rocket has a vernalization requirement. The evidence has suggested that it does not, but repeats of these experiments could be carried out using an identical seed line for each temperature, with a larger number of plants used in each treatment and more biological repeats included for a higher confidence in the real time PCR *DtFLC* expression results. It would also be interesting to include the investigation of other vernalization responsive genes, such as *VIN3*, *VRN1* and *VRN2*, as well as downstream flowering pathway genes such as *DtSOC1* and *DtFT* to try and elucidate

the apparent break in the flowering pathway between the reduced *DtFLC* expression and a reduction in the number of days to bolt.

The photoperiodic requirement of rocket has not been defined. Therefore experiments were set up to test the response of rocket to different photoperiods. Rocket plants were subjected to LD (16 hour light), intermediate day (12 hours light) and SD (8 hour light) conditions and the number of days to initiate bolting was recorded. The experiment was repeated twice, the first with light from fluorescent bulbs for the duration of the photoperiod and the second with fluorescent light for the first 8 hours of the photoperiod, followed by day extensions with incandescent bulbs to make up the remaining photoperiod duration. This was to reduce the growth difference between the photoperiodic treatments. The results showed that rocket was able to flower in all photoperiods but flowered much more quickly under long day conditions. This was true in both experiments, however all plants took longer to flower in experiment 2, despite the experimental set up being the same. Where the incandescent bulbs were used to extend the day length from 8 hours to 12 hours and 16 hours, there was a lengthening in the time it took the plants to initiate. This could be because the day extension incandescent bulbs did not provide the light quantity for optimal growth, but were enough to induce flowering. The discrepancy between the two experiments in the 8 hour photoperiod could have been due to the timing of transplanting or the beginning of the specific photoperiod treatment after germination. Therefore repeating the experiment using the day extension bulbs, would be wise to reduce the discrepancy between the initial two experiments, using a larger number of plants per photoperiod, but also including a diurnal time course for each photoperiod to investigate the expression of diurnally regulated genes, *DtCO*, *DtGI*, *DtSOC1*, *DtFT* and *DtTSF*. Overall the photoperiod experiments in this project have demonstrated that rocket is a facultative LD plant.

Studies in *Arabidopsis* have explored the pattern of ecotype distribution alongside the flowering time response particularly in relation to photoperiod and vernalization responses (Caicedo *et al.*, 2004; Lempe *et al.*, 2005; Stinchcombe *et al.*, 2004). They have linked the different degrees of response to differences in the functionality of *FRI* and *FLC*. This may also explain the vernalization response of rocket, where *DtFLC* expression is lowered, but no change in bolting is observed. With further investigation, particularly of a *FRI* homologue in rocket, a greater understanding of

vernalization in Mediterranean crops could be reached. Nelson *et al.* (2014) investigated the photoperiod responsiveness of summer annual type *B. napus* varieties from Europe and Australia. They found that there was variation in all the varieties screened and the response was closely linked to the ambient temperature floral induction pathway. They also suggested that there was a link between the response to photoperiod and the vernalization requirement of some varieties as the quantitative trait loci (QTL) mapping wasn't as clear as anticipated. This may be the case in rocket so changing the photoperiod length under which the vernalization experiments were conducted could provide further insight into these flowering time pathways. A *Diplotaxis* diversity set could be tested using similar experiments to those outlined in this work for photoperiod and vernalization requirements. This may help determine if there are varieties and species which are more sensitive than the *Dt* WT used here and how the role of genes such as *FLC* and *FRI* may differ between them.

### **7.3. Identification of LB rocket lines from an EMS population**

The aim of this project was to identify rocket plants which are delayed in bolting (i.e. late bolting, LB) so that they could be introduced into breeding programmes to obtain a market leading variety. Currently, in the UK GMOs cannot be utilised so any new varieties must be found using methods which are not deemed GM. EMS has long been used as a method of random mutagenesis followed by selective breeding to identify novel or desired phenotypes in crop plants (Gady *et al.*, 2009; Stephenson *et al.*, 2010; Thurling and Depittayanan, 1992). It was therefore the method of choice for this project in rocket. Rocket seeds of generation SSD4 were treated with two concentrations of EMS, germinated and grew to bolting. Flowers were pollinated using flies to ensure selfed seed was produced. Selection for the late bolting phenotype began at the M<sub>2</sub> generation and candidate plants were backcrossed to wild type to reduce the number of background mutations. The selection of late bolting plants continued until two lines at generation M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub> were chosen. These were segregating into late bolting and wild type bolting plants so genomic DNA was extracted from plants in each line. The DNA from the late bolting plants was pooled together and the DNA from the wild type bolting plants was pooled together. These two sets of pooled DNA were turned into libraries and were sequenced by the Wellcome Trust. The whole genome sequence data was processed

and analysed and a *de novo* assembly constructed. The results revealed that five high quality SNPs with a rank 1 confidence were present between the wild type bolting and late bolting samples, but only three showed any homology to known genes. It is unlikely that this is due to the SNP being mis-called due to the stringency applied to the analysis, but could be due to the SNPs being present in UTR or promoter sequence of a gene, thus enabling it to have an effect on the bolting phenotype, but its location is outside what is currently entered into the NCBI database (<http://www.ncbi.nlm.nih.gov/>). Of the three SNPs, two showed homology to a number of genes with coverage of 37% or lower. This has made interpretation difficult as genes with different functions were identified, so more work is needed to fully understand which genes the SNPs are present in. SNP 3 was found to be homologous to a *B. napus* disease resistance protein TAO1-like when the SNP sequence (Appendix C.1.3) was BLASTed using BLASTn and BLASTx, but also when the contig containing the SNP sequence was BLASTed. There is therefore more confidence in this being the correct gene, however, it is unlikely that this SNP will be contributing to the late bolting phenotype, but it may be linked and so could have co-segregated with the causative SNP. A further 604 SNPs which had a rank of 0.7 or higher were investigated. This revealed one SNP at a rank of 0.8 (SNP 6, Appendix C.1.6) which showed a high homology to a *FT*-like gene and therefore could be linked to the late bolting phenotype. It was found that this SNP was present in intron 1 of the gene and therefore could be involved in regulation of gene expression, perhaps by alternative splicing (Zhang *et al.*, 2011). Further work is needed to fully understand the SNPs identified and which genes they are present in, to be able to sure of their link to the late bolting phenotype. This could be done by further analysis of the sequencing data, but also through verification of the SNPs by PCR and studying the expression of identified genes in both the wild type bolting and late bolting plants.

The two late bolting lines identified in this study need to be selfed for those plants which were late bolting in the segregating population, and the late bolting phenotype selected for. A further self-generation should help increase the number of homozygous alleles in the plants and help create true breeding lines. Another backcross to WT could also be done to reduce the number of non-causative SNPs by 50% (Stephenson *et al.*, 2010) and late bolting plants identified, followed by at least three self-generations to increase the number of homozygous alleles. The robustness

of the late bolting lines could be tested by growing under field conditions and varying the growth temperature in controlled environment cabinets to see if the lines are still late bolting compared to WT under all conditions. This robustness of the phenotype is crucial as growers will expect the lines to be late bolting despite the environmental conditions the plants are cultivated under.

The production of the *de novo* assembly of the whole genome sequence is the first of its kind in *D. tenuifolia* and could be a huge asset to the breeding of rocket. The rocket flowering time homologues were used to BLAST the *de novo* assembly, and matching contigs were found for all of them, although the full coverage of the genes by contigs was not always possible. Other Arabidopsis flowering time genes could be used to search the *de novo* assembly for rocket homologues and this information used to design markers for marker assisted breeding for identification of late bolting lines or for targeted mutagenesis specifically in these genes using CRISPR/Cas9 as exemplified in Chapter 6. This could also be broadened to search for genes known to be involved with other traits, such as disease resistance, to breed better varieties. The drawback with this however, is that the assembly has been constructed using EMS mutagenized plants rather than wild type, so some sequences may contain SNPs due to the EMS treatment. The whole genome sequencing data could also be a resource to design SNP markers, but this also would be limited to the SNPs present for this specific plant line which are associated with the bolting phenotype.

#### **7.4. Targeted mutagenesis of *FT* using CRISPR/Cas9 in rocket and Arabidopsis**

The ability to introduce mutations in a specific location has been around for about 20 years (Bortesi and Fischer, 2015). The technology has rapidly evolved from ZFNs, to TALENs and now CRISPR/Cas9, with each advancement bringing a lower cost, higher specificity and ease of design. CRISPR/Cas9 was the chosen method in this work and collaboration with Professor Yiguo Hong's lab in Hangzhou Normal University allowed several CRISPR/Cas9 constructs to be made targeting common sequences in the *AtFT* and *DtFT* genes. Using floral dipping, these constructs were transformed into Arabidopsis and rocket. Seed was set and threshed for each before selection for transformants began. The T<sub>1</sub> generation were assessed for flowering phenotype, and late flowering Arabidopsis plants were found for FT-1, FT-2 and FT-

96. The DNA was extracted and the *AtFT* gene sequenced for all of these individuals. This showed that plants transformed with the CRISPR/Cas9 construct FT-1 had mutations present in the target site of *AtFT* of either an inserted Adenine or a deletion of an Adenine. One individual showed an insertion of 22 random nucleotides. All the mutations changed the *AtFT* gene such that it could no longer produce a normal functioning FT protein, leading to the change in flowering time phenotype observed. No mutations in the target sequences of FT-2 or FT-96 transformed plants were found, despite the apparent late flowering phenotype. Two rocket plants transformed with CRISPR/Cas9 FT-96 construct were observed to be later bolting than wild type, but sequencing of the *DtFT* to confirm whether a mutation is present or not has not yet been carried out. Good progress has been made in this project with the CRISPR/Cas9 mutagenesis of *DtFT* and *AtFT*, but if the project were to be repeated, then starting this work at an earlier point would be an advantage as a more thorough investigation of the late flowering lines could have been done, carrying through to the T<sub>2</sub> and possible T<sub>3</sub> generation. The mutations *DtFT* in the late bolting rocket lines could also have been found.

Work was done to isolate rocket and Arabidopsis protoplasts and attempts made to transform these cells and regenerate them into whole plants. This was with the aim of providing a method by which the CRISPR/Cas9 constructs could be transiently transformed into the plants without the use of *A. tumefaciens*. The tests showed that protoplasts can be isolated from both species, but further optimisation is needed for rocket as only a small number of isolated protoplasts were viable. This proved a key step in the process as only viable cells could be transformed and regenerated. However, further optimisation is also needed for the transformation to increase the number of cells successfully transformed, and for the regeneration as this was unsuccessful in both species. A method to select for transformed protoplasts would also be an advantage to increase the number of transformed cells entering the regeneration process, so transformed plants are recovered. GFP-tagged plasmids were transformed into Arabidopsis and rocket protoplasts and a FACS machine used to sort the fluorescent cells. However, this only worked for Arabidopsis and gave small numbers of transformed cells which were unsuccessful in the subsequent regeneration. The FACS method could provide a possible way of enriching samples, but the current CRISPR/Cas9 constructs would need to be redesigned to carry a GFP tag, and the process optimised so sorted cells were viable enough to regenerate.

Another method of introducing the CRISPR/Cas9 constructs into rocket could be through virus transfection. Viruses such as Potato Virus X (PVX) and Turnip Mosaic Virus (TuMV) have been successfully used as vectors to introduce foreign genes into plants (Baulcombe *et al.*, 1995; Chapman *et al.*, 1992; Sanchez *et al.*, 1998; Tourino *et al.*, 2008). This technology could be utilised to introduce the CRISPR/Cas9 construct into rocket, providing a virus vector can be found which can infect rocket. The advantage of a virus system is similar to that of transforming and regenerating protoplasts, where the CRISPR/Cas9 would introduce the mutation into the cells and the progeny would carry the mutation, but not the construct, as PVX and TuMV do not move through the germ line.

Once the CRISPR/Cas9 has been optimised for *FT* in rocket where an efficient transformation method has been determined, this technology can then be used for the other key flowering pathway genes to introduce mutations and observe the effect on the bolting phenotype in rocket. It could also be used more widely in other crop species, as long as the sequence for the target gene has been elucidated. This has the potential of producing plants with specific mutations giving desired phenotypes, speeding up breeding programmes and addressing some of the pressing issues in agriculture today.

## 7.5. Conclusions

The aim of this PhD was to produce late bolting rocket lines which can be used to cross into a commercial line where leaf shape, size and taste are uniform across the crop, to create a potential market leading variety. Five objectives were identified in order to give direction to the project so that the aim was achieved.

- Isolate homologues of flowering pathway genes.

Six flowering pathway genes *CO*, *GI*, *FLC*, *SOC1*, *FT* and *TSF* were successfully isolated from rocket. Each was transformed into the equivalent Arabidopsis mutant background and all were shown to complement the mutant flowering phenotype except for *DtGI* which failed to produce transgenic T<sub>1</sub> seed, and *DtTSF* which produced plants which flowered even later than the late flowering *ft-10* mutant. *DtCO*, *DtSOC1* and *DtFT* all restored the flowering phenotype of the late flowering mutant to that of wild type, or gave earlier flowering lines. *DtFLC* restored the

flowering phenotype of the early flowering mutant to that of wild type, or produced later flowering lines. Further analysis of the *Dt* flowering genes was done using developmental and diurnal time course experiments.

- Investigation of vernalization and photoperiodic flowering requirements of rocket

The experiments used to explore these physiological traits of rocket showed that rocket did not have a vernalization requirement as there was no reduction in the number of days to bolt of rocket (seed or plants) vernalized at either 4°C or 10°C. However, a drop in *DtFLC* expression levels was observed after just one week of cold treatment showing that this is a response at the molecular level which is not manifested in a reduced flowering time. The photoperiod experiments showed that rocket is a facultative LD flowering plant as it could flower under all three photoperiod length investigated, but flowered most quickly under a 16 hour photoperiod.

- Create an EMS population of rocket to select for late bolting lines

An EMS population of rocket was created in this project by treating seeds with two different concentrations of EMS. The late bolting phenotype was selected for from the M<sub>2</sub> generation. After one backcross to wild type and selfing the progeny, two late bolting lines were identified for further investigation.

- Use high throughput whole genome sequencing to try to determine the causative SNPs within late lines

The two identified late bolting lines M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub>\_723\_4\_f1\_2 and M<sub>2</sub>\_BC<sub>1</sub>\_S<sub>1</sub>\_723\_4\_f1\_3 had the same M<sub>2</sub> lineage so contained the same EMS mutation. Therefore both lines were sown out and the population segregated into late bolting and wild type bolting plants. DNA was extracted from all the plants and pooled for all the late bolting plants in both lines, and for all the wild type bolting plants in both lines. Two libraries were made, one for each bolting phenotype and sequenced using whole genome sequencing. A sequence analysis pipeline was set up

to clean, quality check and assemble the reads before SNP identification began. Five high confidence (rank 1) SNPs were found to differ between the late bolting and wild type bolting samples, and a sixth SNP (rank 0.8) which was located to a *FT*-like gene, suggesting these could be the mutations responsible for the late bolting phenotype.

- Manipulate rocket flowering pathway genes using CRISPR technology

Following the successful isolation of the rocket flowering time gene homologues, *DtFT* was chosen as the target for manipulation using CRISPR/Cas9. This was because *FT* is a key floral integrator gene where a mutation would change the flowering phenotype of the plant. *DtFT* also has a high sequence homology to *AtFT* providing the ability to test the system in both Arabidopsis and rocket. Four FT CRISPR/Cas9 constructs were made and transformed into Arabidopsis and rocket plants. Late bolting lines were observed in both Arabidopsis and rocket, and CRISPR-mediated mutations in the *AtFT* gene were confirmed by PCR and sequencing of the target region.

Overall, all the objectives were met in the project and novel discoveries made. The aim of the project was achieved as late bolting lines were produced and have been passed to Elsoms Seeds Ltd, the industrial sponsor of the project, for inclusion into their rocket breeding programme to produce market leading varieties. If this project were to be repeated in another crop with similar aims, a larger number of seeds of a pure breeding line should be used in the EMS treatment to increase the likelihood of discovering plants with the desired phenotype. More time and resources should be allocated to the mutagenized population so it could be used in TILLING as well as backcrossed and selfed a number of times for whole genome sequencing as was done in this project. The selection and isolation of flowering time genes in rocket worked well and the approach could be employed in a similar approach for key genes associated with other desired traits. This provided a candidate gene (*DtFT*) for the CRISPR/Cas9 experiments which successfully produced late bolting plants in Arabidopsis and rocket, and could be an approach used for other genes in other crops. This would provide mutagenized plants with the desired phenotype using two

techniques, so market leading varieties could be produced and sold both now and in the future, if the GMO legislation in Europe changes.

## **7.6. Future Work**

The research undertaken in this project is the first of its kind in rocket and therefore provides scope for future work.

The gene isolation work in Chapter 3 has demonstrated that it is possible to find homologues in rocket of Arabidopsis flowering time genes, which were shown to have a high sequence identity and similarity in function and expression patterns. This understanding could lead to the investigation of further important flowering time pathway genes such as *FRI*, *FKF1* and *FCA* using the same techniques, but also could allow other key genes involved in biotic and abiotic stress, pathogen resistance and glucosinolate content to be explored to lead to further potential crop improvements. The incorporation of a rocket diversity set into these investigations could prove useful in providing a source of natural genetic variation which could be crossed into commercial rocket lines to improve traits such as disease resistance and holding ability.

The work carried out in Chapter 4 has provided a valuable insight into some of the physiological traits of rocket. The vernalization experiments have shown that rocket does not have a vernalization requirement but it would be interesting to investigate further the apparent break in the vernalization pathway between the cold response drop in *DtFLC* levels and the lack of response in the flowering time. This could be done by repeating the experiments for more robust data and extended to include other genes thought to be involved in vernalization such as *VIN3*, *VRN1* and *VRN2*, as well as the key flowering time genes, particularly *SOC1* and *FT*. The photoperiod experiments demonstrated that rocket is a facultative long day plant, taking longer to flower under short days compared to long days. This work could be furthered by exploring the expression of key flowering time genes in diurnal time courses in each of the day lengths tested, using both fluorescent and incandescent bulbs as carried out here. Another physiological trait which could be investigated is the time taken to bolt in rocket in different temperatures. The ambient temperature pathway in Arabidopsis has shown to be important in flowering (Fornara *et al.*, 2010; Jarillo and Pineiro, 2011), so similar experiments could be carried out from 16°C to 30°C

conditions. This could provide further insight into ideal conditions for cultivating rocket for high quality produce with delayed bolting.

The EMS population of rocket carried out in Chapter 5 provided two late bolting candidate lines where late bolting and wild type bolting individuals were pooled and sequenced to discover the causative SNPs. Six high confidence SNPs were identified as different between the late and wild type bolting pools and linked to possible genes. This result came at the very end of the project so further work to fully identify the correct genes in which the SNPs lie and their effect on bolting time needs to be done. This could be done by confirmation of the SNPs by PCR and investigating the expression of the SNP carrying genes to see if there is a difference between those in the wild type plants and late bolting plants. Other work could involve finding mutant *Arabidopsis* lines in each candidate gene and screening for flowering time, followed by complementation studies of the rocket genes to see if flowering time is restored. Further breeding of the two late bolting lines used in the whole genome sequencing, as well as other possible lines found in earlier generations to reduce the number of non-causative SNPs as well as to select for those lines with other marketable phenotypes could be carried out by Elsoms Seeds Ltd.

The targeted mutagenesis using CRISPR/Cas9 technology work undertaken in Chapter 6 produced transformed late bolting *Arabidopsis* T<sub>1</sub> plants with mutations in the *FT* gene. Rocket plants were also transformed with the *FT* targeted CRISPR/Cas9 construct and a late bolting phenotype observed in one of these plants. Both plant species should be screened for bolting phenotype and presence of the *FT* mutation in the T<sub>2</sub> generation to provide further evidence of successful mutagenesis. Transforming rocket and *Arabidopsis* protoplasts with CRISPR/Cas9 construct was attempted but not successful, so optimisation of the protoplast isolation and transformation system is needed, followed by the finding of a suitable regeneration method to enable the production of whole plants from the transformed protoplasts. Other methods of transformation of plants with CRISPR/Cas9 constructs such as a virus delivery system should be investigated, and other key flowering time genes targeted to produce the desired late bolting phenotype.

Pursuing all these areas of research could further the understanding of applying knowledge gained from model plants into crops, but also into crop improvement programmes in industry. Continuing the partnership between academia and industry

is vital if this is to be done effectively to ultimately provide new crops and varieties with which to combat some of the issues of global climate change and food security.

### **7.7. Elsoms Seeds Ltd and CASE award**

This project was funded by a BBSRC CASE award with Elsoms Seeds Ltd. This has meant a partnership between our research group and Elsoms Seeds was set up. As well as financially supporting the work, including the purchase of a laptop, contributing to the cost of whole genome sequencing and glasshouse costs, they provided work experience and an insight into how the company works. I spent three months based at the head office in Spalding, where I learnt about the company and their facilities, and took part in the annual open days. I was trained in bud pollination which proved invaluable for producing selfed rocket lines, and in seed quality assessment using germination tests. I was also able to grow and assess some of my rocket experiments there as they have extensive glass and field sites. I was able to contribute to Elsoms breeding programmes by working in their lab on parsnip DNA extraction and marker assisted breeding using SRAP primers and PCR. Elsoms have been able to benefit from this project by including a number of late bolting rocket lines from  $M_3$  and  $M_2\_BC_1\_S_1$  generations into their breeding programme. They will also receive a copy of the *de novo* assembly from the whole genome sequencing to use in SNP marker assisted breeding, as well as a resource for exploring other traits. Elsoms Seeds Ltd has been a great company to be partnered with and is constantly looking to improve their breeding strategy and crop varieties. Rocket is a key crop for them as shown by the collaboration with the University of Warwick with this work, and in pursuing future projects furthering this work, particularly using the CRISPR/Cas9 technology. They also currently have a second rocket PhD project looking at breeding better tasting varieties and understanding the genetics and biochemistry of this, in collaboration with the University of Reading. Together, it is hoped that new, better varieties can be produced, helping them to be market leaders in this crop.

## References

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S. and Terauchi, R.** (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology*, **30**, 174-178.
- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K. and Araki, T.** (2005). FD, a bZIP protein mediating signals from the floral pathway integrator *FT* at the shoot apex. *Science*, **309**, 1052-1056.
- Abou-Elwafa, S. F., Buttner, B., Chia, T., Schulze-Buxloh, G., Hohmann, U., Mutasa-Gottgens, E., Jung, C. and Muller, A. E.** (2011). Conservation and divergence of autonomous pathway genes in the flowering regulatory network of *Beta vulgaris*. *Journal of Experimental Botany*, **62**, 3359-3374.
- Alonso-Blanco, C. and Koornneef, M.** (2000). Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. *Trends in Plant Science*, **5**, 22-29.
- Amasino, R.** (2010). Seasonal and developmental timing of flowering. *Plant Journal*, **61**, 1001-1013.
- An, H. L., Roussot, C., Suarez-Lopez, P., Corbesler, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C. and Coupland, G.** (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. *Development*, **131**, 3615-3626.
- Arias, T. and Pires, J. C.** (2012). A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (*Brassicaceae: Brassicaceae*): Novel clades and potential taxonomic implications. *Taxon*, **61**, 980-988.
- Arisha, M. H., Shah, S. N. M., Gong, Z.-H., Jing, H., Li, C. and Zhang, H.-X.** (2015). Ethyl methane sulfonate induced mutations in M<sub>2</sub> generation and physiological variations in M<sub>1</sub> generation of peppers (*Capsicum annuum* L.). *Frontiers in Plant Science*, **6**, 1-11.
- Arumuganathan, K. and Earle, E. D.** (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reports*, **9**, 208-218.
- Baeurle, I., Smith, L., Baulcombe, D. C. and Dean, C.** (2007). Widespread role for the flowering-time regulators *FCA* and *FPA* in RNA-mediated chromatin silencing. *Science*, **318**, 109-112.
- Baggio, C. and Pimpini, F.** (1995). Preliminary results of agronomic trials on rocket conducted by the ESAV (Agency for the Rural Development of the Veneto Region). *Rocket genetic resources network*, 12-14.

**Balanz, V., Martnez-Fernndez, I. and Ferrndiz, C.** (2014). Sequential action of FRUITFULL as a modulator of the activity of the floral regulators SVP and SOC1. *Journal of Experimental Botany*, **65**, 1193-1203.

**Baulcombe, D. C., Chapman, S. and Cruz, S. S.** (1995). Jellyfish green fluorescent protein as a reporter for virus-infections. *Plant Journal*, **7**, 1045-1053.

**Baumann, K., Venail, J., Berbel, A., Domenech, M. J., Money, T., Conti, L., Hanzawa, Y., Madueno, F. and Bradley, D.** (2015). Changing the spatial pattern of *TFL1* expression reveals its key role in the shoot meristem in controlling Arabidopsis flowering architecture. *Journal of Experimental Botany*, **66**, 4769-4780.

**Bell, L., Oruna-Concha, M. J. and Wagstaff, C.** (2015). Identification and quantification of glucosinolate and flavonol compounds in rocket salad (*Eruca sativa*, *Eruca vesicaria* and *Diplotaxis tenuifolia*) by LC-MS: Highlighting the potential for improving nutritional value of rocket crops. *Food Chemistry*, **172**, 852-861.

**Bennett, R. N., Carvalho, R., Mellon, F. A., Eagles, J. and Rosa, E. A. S.** (2007). Identification and quantification of glucosinolates in sprouts derived from seeds of wild *Eruca sativa* L. (salad rocket) and *Diplotaxis tenuifolia* L. (wild rocket) from diverse geographical locations. *Journal of Agricultural and Food Chemistry*, **55**, 67-74.

**Bennett, R. N., Rosa, E. A. S., Mellon, F. A. and Kroon, P. A.** (2006). Ontogenic profiling of glucosinolates, flavonoids, and other secondary metabolites in *Eruca sativa* (salad rocket), *Diplotaxis erucoides* (wall rocket), *Diplotaxis tenuifolia* (wild rocket), and *Bunias orientalis* (Turkish rocket). *Journal of Agricultural and Food Chemistry*, **54**, 4005-4015.

**Bianco, V. V. and Boari, F.** (1996). *Up-to-date developments on wild rocket cultivation*. In: Rocket: a Mediterranean crop for the world. Legnaro, Italy, 41-.

**Bitinaite, J., Wah, D. A., Aggarwal, A. K. and Schildkraut, I.** (1998). *FokI* dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 10570-10575.

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

**Blazquez, M. A., Green, R., Nilsson, O., Sussman, M. R. and Weigel, D.** (1998). Gibberellins promote flowering of Arabidopsis by activating the *LEAFY* promoter. *Plant Cell*, **10**, 791-800.

**Bluemel, M., Dally, N. and Jung, C.** (2015). Flowering time regulation in crops - what did we learn from Arabidopsis? *Current Opinion in Biotechnology*, **32**, 121-129.

**Bortesi, L. and Fischer, R.** (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, **33**, 41-52.

- Boss, P. K., Bastow, R. M., Mylne, J. S. and Dean, C.** (2004). Multiple pathways in the decision to flower: Enabling, promoting, and resetting. *Plant Cell*, **16**, S18-S31.
- Brambilla, V. and Fornara, F.** (2013). Molecular control of flowering in response to day length in rice. *Journal of Integrative Plant Biology*, **55**, 410-418.
- Brock, R. D.** (1976). Quantitatively inherited variation in *Arabidopsis thaliana* induced by chemical mutagens. *Environmental and Experimental Botany*, **16**, 241-253.
- Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J. and Purugganan, M. D.** (2004). Epistatic interaction between *Arabidopsis* *FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15670-15675.
- Carroll, D.** (2013). Staying on target with CRISPR-Cas. *Nature Biotechnology*, **31**, 807-809.
- Carroll, D., Morton, J. J., Beumer, K. J. and Segal, D. J.** (2006). Design, construction and in vitro testing of zinc finger nucleases. *Nature Protocols*, **1**, 1329-1341.
- Castillejo, C. and Pelaz, S.** (2008). The balance between *CONSTANS* and *TEMPRANILLO* activities determines *FT* expression to trigger flowering. *Current Biology*, **18**, 1338-1343.
- Cavaiuolo, M., Cocetta, G., Bulgari, R., Spinardi, A. and Ferrante, A.** (2015). Identification of innovative potential quality markers in rocket and melon fresh-cut produce. *Food Chemistry*, **188**, 225-233.
- Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J. and Voytas, D. F.** (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, **39**, 7879.
- Chapman, S., Kavanagh, T. and Baulcombe, D.** (1992). Potato virus-x as a vector for gene-expression in plants. *Plant Journal*, **2**, 549-557.
- Chatterjee, G., Sikdar, S. R., Das, S. and Sen, S. K.** (1988). Intergeneric somatic hybrid production through protoplast fusion between *Brassica juncea* and *Diplotaxis muralis*. *Theoretical and Applied Genetics*, **76**, 915-922.
- Cheng, Y. L., Kato, N., Wang, W. M., Li, J. J. and Chen, X. M.** (2003). Two RNA binding proteins, HEN4 and HUM, act in the processing of AGAMOUS Pre-mRNA in *Arabidopsis thaliana*. *Developmental Cell*, **4**, 53-66.
- Chouard, P.** (1960). Vernalization and its relation to dormancy. *Annual Review of Plant Physiology*, **11**, 191-238.

- Christian, M., Qi, Y., Zhang, Y. and Voytas, D. F.** (2013). Targeted Mutagenesis of *Arabidopsis thaliana* Using Engineered TAL Effector Nucleases. *G3-Genes Genomes Genetics*, **3**, 1697-1705.
- Chun, J.-H., Arasu, M. V., Lim, Y.-P. and Kim, S.-J.** (2013). Variation of major glucosinolates in different varieties and lines of rocket salad. *Horticulture, Environment and Biotechnology*, **54**, 206-213.
- Clough, S. J. and Bent, A. F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, **16**, 735-743.
- Cockram, J., Jones, H., Leigh, F. J., O'Sullivan, D., Powell, W., Laurie, D. A. and Greenland, A. J.** (2007). Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *Journal of Experimental Botany*, **58**, 1231-1244.
- Cong, H. Q., Li, Z. Y. and Xu, L.** (2013a). Characterizing developmental and inducible differentiation between juvenile and adult plants of *Aechmea fasciata* treated with ethylene by transcriptomic analysis. *Plant Growth Regulation*, **69**, 247-257.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. and Zhang, F.** (2013b). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, **339**, 819-823.
- Corbesier, L. and Coupland, G.** (2005). Photoperiodic flowering of *Arabidopsis*: integrating genetic and physiological approaches to characterization of the floral stimulus. *Plant Cell and Environment*, **28**, 54-66.
- Curtis, I. S., Nam, H. G., Yun, J. Y. and Seo, K. H.** (2002). Expression of an antisense *GIGANTEA (GI)* gene fragment in transgenic radish causes delayed bolting and flowering. *Transgenic Research*, **11**, 249-256.
- Dahanayake, S. R. and Galwey, N. W.** (1999). Diallel analysis of vernalisation responses in spring rape (*Brassica napus* L.): a basis for adaptation to a Mediterranean environment. *Australian Journal of Agricultural Research*, **50**, 1417-1423.
- DeFrancesco, L.** (2011). Move over ZFNs. *Nature Biotechnology*, **29**, 681-684.
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y. and Yan, N.** (2012). Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science*, **335**, 720-723.
- Devlin, P. F.** (2002). Signs of the time: environmental input to the circadian clock. *Journal of Experimental Botany*, **53**, 1535-1550.

**Domagalska, M. A., Schomburg, F. M., Amasino, R. M., Vierstra, R. D., Nagy, F. and Davis, S. J.** (2007). Attenuation of brassinosteroid signaling enhances *FLC* expression and delays flowering. *Development*, **134**, 2841-2850.

**Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H. and Chandrasegaran, S.** (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Research*, **33**, 5978-5990.

**Edwards, K. D., Anderson, P. E., Hall, A., Salathia, N. S., Locke, J. C. W., Lynn, J. R., Straume, M., Smith, J. Q. and Millar, A. J.** (2006). *FLOWERING LOCUS C* mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. *Plant Cell*, **18**, 639-650.

**Eitas, T. K., Nimchukt, Z. L. and Dangl, J. L.** (2008). *Arabidopsis* TA01 is a TIR-NB-LRR protein that contributes to disease resistance induced by the *Pseudomonas syringae* effector AvrB. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 6475-6480.

**Eschmann-Grupe, G., Neuffer, B. and Hurka, H.** (2004). Extent and structure of genetic variation in two colonising *Diplotaxis* species (*Brassicaceae*) with contrasting breeding systems. *Plant Systematics and Evolution*, **244**, 31-43.

**Esiyok, D.** (1996). *Marketing and Utilization of Rocket in Turkey*. Legnaro, Italy, 86-87. IPGRI.

**Falleh, H., Msilini, N., Oueslati, S., Ksouri, R., Magne, C., Lachaâl, M. and Karray-Bouraoui, N.** (2013). *Diplotaxis harra* and *Diplotaxis simplex* organs: Assessment of phenolics and biological activities before and after fractionation. *Industrial Crops and Products*, **45**, 141-147.

**Fan, C., Hu, R., Zhang, X., Wang, X., Zhang, W., Zhang, Q., Ma, J. and Fu, Y.-F.** (2014). Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean. *BMC Plant Biology*, **14**, 9, 1-14.

**Fausser, F., Schiml, S. and Puchta, H.** (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant Journal*, **79**, 348-359.

**Feng, W., Jacob, Y., Velej, K. M., Ding, L., Yu, X. H., Choe, G. and Michaels, S. D.** (2011). Hypomorphic alleles reveal *FCA*-independent roles for *FY* in the regulation of *FLOWERING LOCUS C*. *Plant Physiology*, **155**, 1425-1434.

**Feng, Z. Y., Zhang, B. T., Ding, W. N., Liu, X. D., Yang, D. L., Wei, P. L., Cao, F. Q., Zhu, S. H., Zhang, F., Mao, Y. F. and Zhu, J. K.** (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, **23**, 1229-1232.

- Fornara, F., de Montaigu, A. and Coupland, G.** (2010). SnapShot: control of flowering in Arabidopsis. *Cell*, **141**, 550-550.
- Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Coupland, G. and Putterill, J.** (1999). *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO Journal*, **18**, 4679-4688.
- Fu, Y. F., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D.** (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, **31**, 822-826.
- Fujimoto, R., Sugimura, T. and Nishio, T.** (2006). Gene conversion from SLG to SRK resulting in self-compatibility in *Brassica rapa*. *FEBS Letters*, **580**, 425-430.
- Gady, A. L. F., Hermans, F. W. K., Van de Wal, M., van Loo, E. N., Visser, R. G. F. and Bachem, C. W. B.** (2009). Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations. *Plant Methods*, **5**, 13, 1-14.
- Galvao, V. C., Horrer, D., Kuettner, F. and Schmid, M.** (2012). Spatial control of flowering by DELLA proteins in *Arabidopsis thaliana*. *Development*, **139**, 4072-4082.
- Garg, H., Atri, C., Sandhu, P. S., Kaur, B., Renton, M., Banga, S. K., Singh, H., Singh, C., Barbetti, M. J. and Banga, S. S.** (2010). High level of resistance to *Sclerotinia sclerotiorum* in introgression lines derived from hybridization between wild crucifers and the crop *Brassica* species *B. napus* and *B. juncea*. *Field Crops Research*, **117**, 51-58.
- Gendall, A. R., Levy, Y. Y., Wilson, A. and Dean, C.** (2001). The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*, **107**, 525-535.
- Gill, M.** (2008). *Wild Rocket Sales on the Up*. In: The Vegetable Farmer. January, 27-28. ACT Publishing.
- Gresta, F., Cristaudo, A., Onofri, A., Restuccia, A. and Avola, G.** (2010). Germination response of four pasture species to temperature, light, and post-harvest period. *Plant Biosystems*, **144**, 849-856.
- Griffiths, J., Murase, K., Rieu, I., Zentella, R., Zhang, Z.-L., Powers, S. J., Gong, F., Phillips, A. L., Hedden, P., Sun, T.-p. and Thomas, S. G.** (2006). Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. *Plant Cell*, **18**, 3399-3414.
- Grillo, O., Draper, D., Venora, G. and Bautista Martinez-Laborde, J.** (2012). Seed image analysis and taxonomy of *Diplotaxis* DC. (*Brassicaceae*, *Brassicaceae*). *Systematics and Biodiversity*, **10**, 57-70.

- Guo, D. P., Shah, G. A., Zeng, G. W. and Zheng, S. J.** (2004). The interaction of plant growth regulators and vernalization on the growth and flowering of cauliflower (*Brassica oleracea* var. botrytis). *Plant Growth Regulation*, **43**, 163-171.
- Guo, Y., Hans, H., Christian, J. and Molina, C.** (2014). Mutations in single *FT*- and *TFL1*-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. *Frontiers in Plant Science*, **5**, 282, 1-12.
- Hackett, W. P. and Hartmann, H. T.** (1967). The influence of temperature on floral initiation in the olive. *Physiologia Plantarum*, **20**, 430-436.
- Hall, M. K. D., Jobling, J. J. and Rogers, G. S.** (2015). Variations in the most abundant types of glucosinolates found in the leaves of baby leaf rocket under typical commercial conditions. *Journal of the Science of Food and Agriculture*, **95**, 552-559.
- Hanano, S. and Goto, K.** (2011). Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. *Plant Cell*, **23**, 3172-3184.
- Harberd, D.** (1976). Cytotaxonomic studies of *Brassica* and related genera. In *The biology and chemistry of the Cruciferae*, pp. 47-68. Edited by J. Vaughan, A. MacLeod and B. Jones. London: Academic.
- Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H. and Huijser, P.** (2000). Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. *Plant Journal*, **21**, 351-360.
- Hayama, R., Izawa, T. and Shimamoto, K.** (2002). Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. *Plant and Cell Physiology*, **43**, 494-504.
- Hayama, R., Yokoi, S., Tamaki, S., Yano, M. and Shimamoto, K.** (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature*, **422**, 719-722.
- He, Y. H., Michaels, S. D. and Amasino, R. M.** (2003). Regulation of flowering time by histone acetylation in Arabidopsis. *Science*, **302**, 1751-1754.
- Hedden, P. and Phillips, A. L.** (2000). Gibberellin metabolism: new insights revealed by the genes. *Trends in Plant Science*, **5**, 523-530.
- Helliwell, C. A., Wood, C. C., Robertson, M., Peacock, W. J. and Dennis, E. S.** (2006). The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *Plant Journal*, **46**, 183-192.

- Hornyik, C., Terzi, L. C. and Simpson, G. G.** (2010). The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Developmental Cell*, **18**, 203-213.
- Hou, J., Long, Y., Raman, H., Zou, X., Wang, J., Dai, S., Xiao, Q., Li, C., Fan, L., Liu, B. and Meng, J.** (2012). A Tourist-like MITE insertion in the upstream region of the *BnFLC.A10* gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.). *BMC Plant Biology*, **12**, 238, 1-13.
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y. Q., Fine, E. J., Wu, X. B., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. and Zhang, F.** (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, **31**, 827-834.
- Huq, E., Tepperman, J. M. and Quail, P. H.** (2000). GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9789-9794.
- Immink, R. G. H., Pose, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F. L., de Folter, S., van der Wal, F., van Dijk, A. D. J., Schmid, M. and Angenent, G. C.** (2012). Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. *Plant Physiology*, **160**, 433-449.
- Iqbal, M., Navabi, A., Yang, R.-C., Salmon, D. F. and Spaner, D.** (2007). Molecular characterization of vernalization response genes in Canadian spring wheat. *Genome*, **50**, 511-516.
- Ito, S., Song, Y. H., Josephson-Day, A. R., Miller, R. J., Breton, G., Olmstead, R. G. and Imaizumi, T.** (2012). FLOWERING BHLH transcriptional activators control expression of the photoperiodic flowering regulator *CONSTANS* in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3582-3587.
- Izawa, T., Oikawa, T., Sugiyama, N., Tanisaka, T., Yano, M. and Shimamoto, K.** (2002). Phytochrome mediates the external light signal to repress *FT* orthologs in photoperiodic flowering of rice. *Genes & Development*, **16**, 2006-2020.
- Jarillo, J. A. and Pineiro, M.** (2011). Timing is everything in plant development. The central role of floral repressors. *Plant Science*, **181**, 364-378.
- Jiang, W., Yang, B. and Weeks, D. P.** (2014). Efficient CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana* and inheritance of modified genes in the T<sub>2</sub> and T<sub>3</sub> generations. *PLOS One*, **9**, 6, 1-10.
- Jiang, W. Z., Zhou, H. B., Bi, H. H., Fromm, M., Yang, B. and Weeks, D. P.** (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Research*, **41**, 12.

- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E.** (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
- Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. and Dean, C.** (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science*, **290**, 344-347.
- Jung, C. and Mueller, A. E.** (2009). Flowering time control and applications in plant breeding. *Trends in Plant Science*, **14**, 563-573.
- Kardailsky, I., Shukla, V. K., Ahn, J. H., Dagenais, N., Christensen, S. K., Nguyen, J. T., Chory, J., Harrison, M. J. and Weigel, D.** (1999). Activation tagging of the floral inducer FT. *Science*, **286**, 1962-1965.
- Karimi, M., Inze, D. and Depicker, A.** (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science*, **7**, 193-195.
- Kenigsbuch, D., Ovadia, A., Ivanova, Y., Chalupowicz, D., Maurer, D., Aharon, Z., Vinokur, Y. and Aharoni, N.** (2009). *Wild rocket* (*Diplotaxis tenuifolia*) mutant with late flowering and delay in postharvest senescence. In: IV<sup>th</sup> Balkan Symposium on Vegetables and Potatoes. 830, 91-96. Acta Horticultura.
- Kenigsbuch, D., Ovadia, A., Shahar-Ivanova, Y., Chalupowicz, D. and Maurer, D.** (2014). "Rock-Ad", a new wild rocket (*Diplotaxis tenuifolia*) mutant with late flowering and delayed postharvest senescence. *Scientia Horticulturae*, **174**, 17-23.
- Kim, J. J., Lee, J. H., Kim, W., Jung, H. S., Huijser, P. and Ahn, J. H.** (2012). The microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3 module regulates ambient temperature-responsive flowering via *FLOWERING LOCUS T* in Arabidopsis. *Plant Physiology*, **159**, 461-478.
- Kim, K. W., Shin, J. H., Moon, J., Kim, M., Lee, J., Park, M. C. and Lee, I.** (2003). The function of the flowering time gene *AGL20* is conserved in crucifers. *Molecules and Cells*, **16**, 136-141.
- Kim, S., Choi, K., Park, C., Hwang, H.-J. and Lee, I.** (2006a). SUPPRESSOR OFFRIGIDA4, encoding a C2H2-type zinc finger protein, represses flowering by transcriptional activation of Arabidopsis *FLOWERING LOCUS C*. *Plant Cell*, **18**, 2985-2998.
- Kim, S.-Y., Park, B.-S., Kwon, S.-J., Kim, J., Lim, M.-H., Park, Y.-D., Kim, D. Y., Suh, S.-C., Jin, Y.-M., Ahn, J. H. and Lee, Y.-H.** (2007). Delayed flowering time in Arabidopsis and *Brassica rapa* by the overexpression of *FLOWERING LOCUS C* (*FLC*) homologs isolated from Chinese cabbage (*Brassica rapa* L. ssp *pekinensis*). *Plant Cell Reports*, **26**, 327-336.
- Kim, S. Y., Yu, X. and Michaels, S. D.** (2008). Regulation of *CONSTANS* and *FLOWERING LOCUS T* expression in response to changing light quality. *Plant Physiology*, **148**, 269-279.

**Kim, W., Park, T. I., Yoo, S. J., Jun, A. R. and Ahn, J. H.** (2013). Generation and analysis of a complete mutant set for the Arabidopsis FT/TFL1 family shows specific effects on thermo-sensitive flowering regulation. *Journal of Experimental Botany*, **64**, 1715-1729.

**Kim, Y., Schumaker, K. S. and Zhu, J.-K.** (2006b). EMS Mutagenesis of *Arabidopsis*. In *Methods in Molecular Biology*, pp. 101-103. Edited by J. Salinas and J. J. Sanchez-Serrano. Totowa, NJ: Humana Press Inc.

**Klee, H. J., Muskopf, Y. M. and Gasser, C. S.** (1987). Cloning of an *Arabidopsis thaliana* gene encoding 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE - sequence-analysis and manipulation to obtain glyphosate-tolerant plants. *Molecular & General Genetics*, **210**, 437-442.

**Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. and Araki, T.** (1999). A pair of related genes with antagonistic roles in mediating flowering signals. *Science*, **286**, 1960-1962.

**Lai, Y. P., Huang, J., Wang, L. F., Li, J. and Wu, Z. R.** (2004). A new approach to random mutagenesis in vitro. *Biotechnology and Bioengineering*, **86**, 622-627.

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

**Lee, H., Yoo, S. J., Lee, J. H., Kim, W., Yoo, S. K., Fitzgerald, H., Carrington, J. C. and Ahn, J. H.** (2010). Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. *Nucleic Acids Research*, **38**, 3081-3093.

**Lee, R., Baldwin, S., Kenel, F., McCallum, J. and Macknight, R.** (2013). FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nature Communications*, **4**, 2884, 1-9.

**Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M. and Weigel, D.** (2005). Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLOS Genetics*, **1**, 109-118.

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

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

**Li, D. L., Qiu, Z. W., Shao, Y. J., Chen, Y. T., Guan, Y. T., Liu, M. Z., Li, Y. M., Gao, N., Wang, L. R., Lu, X. L., Zhao, Y. X. and Liu, M. Y.** (2013a). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nature Biotechnology*, **31**, 681-683.

- Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D. D., Bush, J., Church, G. M. and Sheen, J.** (2013b). Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnology*, **31**, 688-691.
- Liang, Z., Zhang, K., Chen, K. and Gao, C.** (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *Journal of Genetics and Genomics*, **41**, 63-68.
- Lim, M. H., Kim, J., Kim, Y. S., Chung, K. S., Seo, Y. H., Lee, I., Hong, C. B., Kim, H. J. and Park, C. M.** (2004). A new *Arabidopsis* gene, *FLK*, encodes an RNA binding protein with K homology motifs and regulates flowering time via *FLOWERING LOCUS C*. *Plant Cell*, **16**, 731-740.
- Lin, S. I., Wang, J. G., Poon, S. Y., Su, C. L., Wang, S. S. and Chiou, T. J.** (2005). Differential regulation of expression by vernalization *FLOWERING LOCUS C* in cabbage and *Arabidopsis*. *Plant Physiology*, **137**, 1037-1048.
- Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T. and Yu, H.** (2007). Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. *Development*, **134**, 1901-1910.
- Liu, L., Liu, C., Hou, X., Xi, W., Shen, L., Tao, Z., Wang, Y. and Yu, H.** (2012). *FTIP1* is an essential regulator required for florigen transport. *PLOS Biology*, **10**, 4, 1-14.
- Logemann, J., Schell, J. and Willmitzer, L.** (1987). Improved method for the isolation of RNA from plant-tissues. *Analytical Biochemistry*, **163**, 16-20.
- Lokke, M. M., Seefeldt, H. F. and Edelenbos, M.** (2012). Freshness and sensory quality of packaged wild rocket. *Postharvest Biology and Technology*, **73**, 99-106.
- Lunn, J. E., Feil, R., Hendriks, J. H. M., Gibon, Y., Morcuende, R., Osuna, D., Scheible, W. R., Carillo, P., Hajirezaei, M. R. and Stitt, M.** (2006). Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in *Arabidopsis thaliana*. *Biochemical Journal*, **397**, 139-148.
- Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. and Dean, C.** (1997). *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell*, **89**, 737-745.
- Mahfouz, M. M., Li, L. X., Shamimuzzaman, M., Wibowo, A., Fang, X. Y. and Zhu, J. K.** (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 2623-2628.
- Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. H. and Church, G. M.** (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, **31**, 833-838.

**Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J.-K.** (2013). Application of the CRISPR/Cas system for efficient genome engineering in plants. *Molecular Plant*, **6**, 2008-2011.

**Martin-Tryon, E. L., Kreps, J. A. and Harmer, S. L.** (2007). GIGANTEA acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation. *Plant Physiology*, **143**, 473-486.

**Marton, I., Zuker, A., Shklarman, E., Zeevi, V., Tovkach, A., Roffe, S., Ovadis, M., Tzfira, T. and Vainstein, A.** (2010). Nontransgenic genome modification in plant cells. *Plant Physiology*, **154**, 1079-1087.

**Mathieu, J., Yant, L. J., Muerdter, F., Kuettner, F. and Schmid, M.** (2009). Repression of flowering by the miR172 target SMZ. *PLOS Biology*, **7**, 7, 1-15.

**Matsoukas, I. G., Massiah, A. J. and Thomas, B.** (2012). Florigenic and antiflorigenic signaling in plants. *Plant and Cell Physiology*, **53**, 1827-1842.

**Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A. and Beeckman, T.** (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nature Genetics*, **40**, 1489-1492.

**Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H. and Qu, L.-J.** (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Research*, **23**, 1233-1236.

**Michaels, S. D.** (2009). Flowering time regulation produces much fruit. *Current Opinion in Plant Biology*, **12**, 75-80.

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

**Michaels, S. D. and Amasino, R. M.** (2000). Memories of winter: vernalization and the competence to flower. *Plant Cell and Environment*, **23**, 1145-1153.

**Michaels, S. D., He, Y. H., Scortecci, K. C. and Amasino, R. M.** (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10102-10107.

**Miller, T. A., Muslin, E. H. and Dorweiler, J. E.** (2008). A maize *CONSTANS-like* gene, *CONZ1*, exhibits distinct diurnal expression patterns in varied photoperiods. *Planta*, **227**, 1377-1388.

**Mimida, N., Goto, K., Kobayashi, Y., Araki, T., Ahn, J. H., Weigel, D., Murata, M., Motoyoshi, F. and Sakamoto, W.** (2001). Functional divergence of the *TFL1-like* gene family in Arabidopsis revealed by characterization of a novel homologue. *Genes to Cells*, **6**, 327-336.

**Mishra, P. and Panigrahi, K. C.** (2015). GIGANTEA - an emerging story. *Frontiers in Plant Science*, **6**, 8, 1-15.

**Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J. and Coupland, G.** (2005). Distinct roles of *GIGANTEA* in promoting flowering and regulating circadian rhythms in Arabidopsis. *Plant Cell*, **17**, 2255-2270.

**Mohamedien, S.** (1994). *Rocket cultivation in Egypt*. In: Rocket Genetic Resources Network. Lisbon, Portugal, 61-62. IPGRI.

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

**Morbitzer, R., Romer, P., Boch, J. and Lahaye, T.** (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 21617-21622.

**Nakamichi, N.** (2015). Adaptation to the local environment by modifications of the photoperiod response in crops. *Plant and Cell Physiology*, **56**, 594-604.

**Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G. and Kamoun, S.** (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology*, **31**, 691-693.

**Nelson, M. N., Rajasekaran, R., Smith, A., Chen, S., Beeck, C. P., Siddique, K. H. M. and Cowling, W. A.** (2014). Quantitative trait loci for thermal time to flowering and photoperiod responsiveness discovered in summer annual-type *Brassica napus* L. *PLOS One*, **9**, 7, 1-11.

**Nordborg, M. and 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**, 470-475.

**Park, D. H., Somers, D. E., Kim, Y. S., Choy, Y. H., Lim, H. K., Soh, M. S., Kim, H. J., Kay, S. A. and Nam, H. G.** (1999). Control of circadian rhythms and photoperiodic flowering by the Arabidopsis *GIGANTEA* gene. *Science*, **285**, 1579-1582.

**Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E. B., Doudna, J. A. and Liu, D. R.** (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, **31**, 839-845.

**Patterson, H. D., Williams, E. R. and Hunter, E. A.** (1978). Block designs for variety trials. *Journal of Agricultural Science*, **90**, 395-400.

**Pavletich, N. P. and Pabo, C. O.** (1991). Zinc finger DNA recognition - crystal-structure of a ZIF268-DNA complex at 2.1-Å. *Science*, **252**, 809-817.

**Perilleux, C., Pieltain, A., Jacquemin, G., Bouche, F., Detry, N., D'Aloia, M., Thiry, L., Aljochim, P., Delansnay, M., Mathieu, A.-S., Lutts, S. and Tocquin, P.** (2013). A root chicory MADS box sequence and the Arabidopsis flowering repressor FLC share common features that suggest conserved function in vernalization and de-vernalization responses. *Plant Journal*, **75**, 390-402.

**Phinney, B.** (1983). The history of gibberellins. In *The Biochemistry and Physiology of Gibberellins*, pp. 19-52. Edited by A. Crozier. New York, USA: Praeger Publishers.

**Pignone, D. and Martinez-Laborde, J.** (2011). Diplotaxis. In *Wild Crop Relatives: Genomic and Breeding Resources; Oilseeds*, pp. 137-147. Edited by C. Kole: Springer.

**Pimpini, F. and Enzo, M.** (1996). *Present status and prospects for rocket cultivation in the Veneto region*. In: Rocket: a Mediterranean crop for the world. Legnaro, Italy, 51-. IPGRI.

**Pin, P. A., Benloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J. J. L. and Nilsson, O.** (2010). An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. *Science*, **330**, 1397-1400.

**Rakow, G., Edwards, D., Batley, J., Parkin, I. and Kole, C.** (2011). Classical genetics and traditional breeding. In *Genetics, genomics and breeding of oilseed brassicas*, pp. 73-84. Edited by: Science Publishers, Inc.

**Rantanen, M., Kurokura, T., Mouhu, K., Pinho, P., Tetri, E., Halonen, L., Palonen, P., Elomaa, P. and Hytonen, T.** (2014). Light quality regulates flowering in *FvFT1/FvTFL1* dependent manner in the woodland strawberry *Fragaria vesca*. *Frontiers in Plant Science*, **5**, 271, 1-11.

**Reeves, P. A., He, Y., Schmitz, R. J., Amasino, R. M., Panella, L. W. and Richards, C. M.** (2007). Evolutionary conservation of the FLOWERING LOCUS C-mediated vernalization response: Evidence from the sugar beet (*Beta vulgaris*). *Genetics*, **176**, 295-307.

**Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D. and Joung, J. K.** (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology*, **30**, 460-465.

**Robert, L. S., Robson, F., Sharpe, A., Lydiate, D. and Coupland, G.** (1998). Conserved structure and function of the Arabidopsis flowering time gene CONSTANS in *Brassica napus*. *Plant Molecular Biology*, **37**, 763-772.

**Robson, F., Costa, M. M. R., Hepworth, S. R., Vizir, I., Pineiro, M., Reeves, P. H., Putterill, J. and Coupland, G.** (2001). Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. *Plant Journal*, **28**, 619-631.

**Rodriguez, L., Gonzalez-Guzman, M., Diaz, M., Rodrigues, A., Izquierdo-Garcia, A. C., Peirats-Llobet, M., Fernandez, M. A., Antoni, R., Fernandez, D., Marquez, J. A., Mulet, J. M., Albert, A. and Rodriguez, P. L.** (2014). C2-Domain abscisic acid-related proteins mediate the interaction of PYR/PYL/RCAR abscisic acid receptors with the plasma membrane and regulate abscisic acid sensitivity in *Arabidopsis*. *Plant Cell*, **26**, 4802-4820.

**Rodriguez-Falcon, M., Bou, J. and Prat, S.** (2006). Seasonal control of tuberization in potato: Conserved elements with the flowering response. *Annual Review of Plant Biology*, **57**, 151-180.

**Roemer, P., Hahn, S., Jordan, T., Strauss, T., Bonas, U. and Lahaye, T.** (2007). Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science*, **318**, 645-648.

**Runkle, E. S., Padhye, S. R., Oh, W. and Getter, K.** (2012). Replacing incandescent lamps with compact fluorescent lamps may delay flowering. *Scientia Horticulturae*, **143**, 56-61.

**Samach, A., Onouchi, H., Gold, S. E., Ditta, G. S., Schwarz-Sommer, Z., Yanofsky, M. F. and Coupland, G.** (2000). Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science*, **288**, 1613-1616.

**Sanchez, F., Martinez-Herrera, D., Aguilar, I. and Ponz, F.** (1998). Infectivity of turnip mosaic potyvirus cDNA clones and transcripts on the systemic host *Arabidopsis thaliana* and local lesion hosts. *Virus Research*, **55**, 207-219.

**Sanchez-Bermejo, E., Mendez-Vigo, B., Xavier Pico, F., Martinez-Zapater, J. M. and Alonso-Blanco, C.** (2012). Novel natural alleles at *FLC* and *LVR* loci account for enhanced vernalization responses in *Arabidopsis thaliana*. *Plant Cell and Environment*, **35**, 1672-1684.

**Sawa, M. and Kay, S. A.** (2011). *GIGANTEA* directly activates *FLOWERING LOCUS T* in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11698-11703.

**Sawa, M., Nusinow, D. A., Kay, S. A. and Imaizumi, T.** (2007). FKF1 and *GIGANTEA* complex formation is required for day-length measurement in *Arabidopsis*. *Science*, **318**, 261-265.

**Sayers, E., Barrett, T., Benson, D., Bolton, E., Bryant, S., Canese, K., Chetvernin, V., Church, D., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I., Geer, L., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D., Lu, Z., Madden, T., Madej, T., Maglott, D., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K., Schuler, G., Sequeira, E., Sherry, S., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G.,**

**Tatusova, T., Wagner, L., Wang, Y., Wilbur, W., Yaschenko, E. and Ye, J.** (2011). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **39**, 38-51.

**Sessa, R. A., Bennett, M. H., Lewis, M. J., Mansfield, J. W. and Beale, M. H.** (2000). Metabolite profiling of sesquiterpene lactones from *Lactuca* species - Major latex components are novel oxalate and sulfate conjugates of lactucin and its derivatives. *Journal of Biological Chemistry*, **275**, 26877-26884.

**Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D. F., Zheng, X., Zhang, Y. and Gao, C.** (2013a). Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Molecular Plant*, **6**, 1365-1368.

**Shan, Q. W., Wang, Y. P., Li, J., Zhang, Y., Chen, K. L., Liang, Z., Zhang, K., Liu, J. X., Xi, J. J., Qiu, J. L. and Gao, C. X.** (2013b). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, **31**, 686-688.

**Shen, C., Xu, Y., Huang, J., Wang, Z., Qiu, J. and Huang, Y.** (2014). Molecular characterization and expression analysis of the critical floral genes in hickory (*Carya cathayensis* Sarg.). *Plant Physiology and Biochemistry*, **83**, 142-150.

**Shrestha, R., Gomez-Ariza, J., Brambilla, V. and Fornara, F.** (2014). Molecular control of seasonal flowering in rice, Arabidopsis and temperate cereals. *Annals of Botany*, **114**, 1445-1458.

**Shukla, V. K., Doyon, Y., Miller, J. C., DeKolver, R. C., Moehle, E. A., Worden, S. E., Mitchell, J. C., Arnold, N. L., Gopalan, S., Meng, X., Choi, V. M., Rock, J. M., Wu, Y.-Y., Katibah, G. E., Zhifang, G., McCaskill, D., Simpson, M. A., Blakeslee, B., Greenwalt, S. A., Butler, H. J., Hinkley, S. J., Zhang, L., Rebar, E. J., Gregory, P. D. and Urnov, F. D.** (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature*, **459**, 437-443.

**Sikdar, S. R., Chatterjee, G., Das, S. and Sen, S. K.** (1987). Regeneration of plants from mesophyll protoplasts of the wild crucifer *Eruca sativa* Lam. *Plant Cell Reports*, **6**, 486-489.

**Sikdar, S. R., Sengupta, S., Das, S. and Sen, S. K.** (1990). Plant-regeneration from mesophyll protoplasts of *Diplotaxis muralis*, a wild crucifer. *Plant Cell Reports*, **8**, 722-725.

**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**, 570-574.

**Simpson, G. G., Dijkwel, P. P., Quesada, V., Henderson, I. and Dean, C.** (2003). FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition. *Cell*, **113**, 777-787.

- Song, J., Angel, A., Howard, M. and Dean, C.** (2012). Vernalization - a cold-induced epigenetic switch. *Journal of Cell Science*, **125**, 3723-3731.
- Song, Y. H., Shim, J. S., Kinmonth-Schultz, H. A. and Imaizumi, T.** (2015). Photoperiodic flowering: time measurement mechanisms in leaves. *Annual Review of Plant Biology*, **66**, 441-464.
- Srikanth, A. and Schmid, M.** (2011). Regulation of flowering time: all roads lead to Rome. *Cellular and Molecular Life Sciences*, **68**, 2013-2037.
- Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G. J. and Ostergaard, L.** (2010). A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biology*, **10**, 62, 1-10.
- Stewart, C. N. and Via, L. E.** (1993). A rapid CTAB DNA isolation technique useful for rapid fingerprinting and other PCR applications. *Biotechniques*, **14**, 748-751.
- Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., Purugganan, M. D. and Schmitt, J.** (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 4712-4717.
- Strange, A., Li, P., Lister, C., Anderson, J., Warthmann, N., Shindo, C., Irwin, J., Nordborg, M. and Dean, C.** (2011). Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions. *PLOS One*, **6**, 5, 1-11.
- Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F. and Coupland, G.** (2001). *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*, **410**, 1116-1120.
- Sugisaki, H. and Kanazawa, S.** (1981). New restriction endonucleases from *Flavobacterium okeanokoites* (*FokI*) and *Micrococcus luteus* (*MluI*). *Gene*, **16**, 73-78.
- Sung, S. B. and Amasino, R. M.** (2004). Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature*, **427**, 159-164.
- Swarup, S., Yang, Y. N., Kingsley, M. T. and Gabriel, D. W.** (1992). An *Xanthomonas citri* pathogenicity gene, *PTHA*, pleiotropically encodes gratuitous avirulence on nonhosts. *Molecular Plant-Microbe Interactions*, **5**, 204-213.
- Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, E. S. and Peacock, W. J.** (2001). Control of flowering time by FLC orthologues in *Brassica napus*. *Plant Journal*, **28**, 545-553.

**Tao, Z., Shen, L. S., Liu, C., Liu, L., Yan, Y. Y. and Yu, H.** (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. *Plant Journal*, **70**, 549-561.

**Thimijan, R. W. and Heins, R. D.** (1983). Photometric, radiometric, and quantum light units of measure - a review of procedures for interconversion. *Hortscience*, **18**, 818-822.

**Thomas, B.** (2006). Light signals and flowering. *Journal of Experimental Botany*, **57**, 3387-3393.

**Thomas, B., Carre, I. and Jackson, S.** (2006). Photoperiodism and Flowering. *Molecular Biology and Biotechnology of Flowering, 2nd Edition*, 3-25.

**Thurling, N. and Depittayanan, V.** (1992). EMS induction of early flowering mutants in spring rape (*Brassica napus*). *Plant Breeding*, **108**, 177-184.

**Till, B. J., Zerr, T., Comai, L. and Henikoff, S.** (2006). A protocol for TILLING and Ecotilling in plants and animals. *Nature Protocols*, **1**, 2465-2477.

**Tilman, D., Balzer, C., Hill, J. and Befort, B. L.** (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 20260-20264.

**Tourino, A., Sanchez, E., Fereres, A. and Ponz, F.** (2008). High expression of foreign proteins from a biosafe viral vector derived from Turnip mosaic virus. *Spanish Journal of Agricultural Research*, **6**, 48-58.

**Townsend, J. A., Wright, D. A., Winfrey, R. J., Fu, F. L., Maeder, M. L., Joung, J. K. and Voytas, D. F.** (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, **459**, 442-446.

**Tsai, H., Missirian, V., Ngo, K. J., Tran, R. K., Chan, S. R., Sundaresan, V. and Comai, L.** (2013). Production of a high-efficiency TILLING population through polyploidization. *Plant Physiology*, **161**, 1604-1614.

**Turner, A., Beales, J., Faure, S., Dunford, R. P. and Laurie, D. A.** (2005). The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science*, **310**, 1031-1034.

**Upadhyay, S. K., Kumar, J., Alok, A. and Tuli, R.** (2013). RNA-guided genome editing for target gene mutations in wheat. *G3-Genes Genomes Genetics*, **3**, 2233-2238.

**Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A. and Coupland, G.** (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science*, **303**, 1003-1006.

- Velasquez, N., Murphy, J. F. and Suh, S.-J.** (2012). Electroporetic transfection of pepper protoplasts with plant potyviruses. *Journal of Virological Methods*, **179**, 154-160.
- Vogt, S. H., Weyens, G., Lefebvre, M., Bork, B., Schechert, A. and Mueller, A. E.** (2014). The *FLC-like* gene *BvFL1* is not a major regulator of vernalization response in biennial beets. *Frontiers in Plant Science*, **5**, 146, 1-12.
- Wahl, V., Ponnu, J., Schlereth, A., Arrivault, S., Langenecker, T., Franke, A., Feil, R., Lunn, J. E., Stitt, M. and Schmid, M.** (2013). Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science*, **339**, 704-707.
- Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. and Jaenisch, R.** (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910-918.
- Wang, J., Hopkins, C. J., Hou, J. N., Zou, X. X., Wang, C. N., Long, Y., Kurup, S., King, G. J. and Meng, J. L.** (2012). Promoter variation and transcript divergence in *Brassicaceae* lineages of *FLOWERING LOCUS T*. *PLOS One*, **7**, 10, 1-11.
- Wang, J. W.** (2014). Regulation of flowering time by the miR156-mediated age pathway. *Journal of Experimental Botany*, **65**, 4723-4730.
- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.-L.** (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, **32**, 947-951.
- Warwick, S. I. and Al-Shehbaz, I. A.** (2006). Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution*, **259**, 237-248.
- Warwick, S. I., Black, L. D. and Aguinagalde, I.** (1992). Molecular systematics of *Brassica* and allied genera (subtribe brassicinae, *brassicaceae*) - chloroplast DNA variation in the genus *Diplotaxis*. *Theoretical and Applied Genetics*, **83**, 839-850.
- Weigel, D.** (1995). The genetics of flower development: From floral induction to ovule morphogenesis. *Annual Review of Genetics*, **29**, 19-39.
- Weightman, R. M., Huckle, A. J., Roques, S. E., Ginsburg, D. and Dyer, C. J.** (2012). Factors influencing tissue nitrate concentration in field-grown wild rocket (*Diplotaxis tenuifolia*) in southern England. *Food Additives and Contaminants Part a-Chemistry Analysis Control Exposure & Risk Assessment*, **29**, 1425-1435.
- Weis, C., Hueckelhoven, R. and Eichmann, R.** (2013). LIFEGUARD proteins support plant colonization by biotrophic powdery mildew fungi. *Journal of Experimental Botany*, **64**, 3855-3867.

**Wellensiek, S. J.** (1964). Dividing cells as the prerequisite for vernalization. *Plant physiology*, **39**, 832-5.

**Wickland, D. P. and Hanzawa, Y.** (2015). The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms. *Molecular Plant*, **8**, 983-987

**Wu, J., Wei, K., Cheng, F., Li, S., Wang, Q., Zhao, J., Bonnema, G. and Wang, X.** (2012). A naturally occurring InDel variation in *BraA.FLC.b(BrFLC2)* associated with flowering time variation in *Brassica rapa*. *BMC Plant Biology*, **12**, 151, 1-9.

**Xiao, A., Wu, Y. D., Yang, Z. P., Hu, Y. Y., Wang, W. Y., Zhang, Y. T., Kong, L., Gao, G., Zhu, Z. Y., Lin, S. and Zhang, B.** (2013). EENdb: a database and knowledge base of ZFNs and TALENs for endonuclease engineering. *Nucleic Acids Research*, **41**, D415-D422.

**Xie, Q., Lou, P., Hermand, V., Aman, R., Park, H. J., Yun, D.-J., Kim, W. Y., Salmela, M. J., Ewers, B. E., Weinig, C., Khan, S. L., Schaible, D. L. P. and McClung, C. R.** (2015). Allelic polymorphism of GIGANTEA is responsible for naturally occurring variation in circadian period in *Brassica rapa*. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 3829-3834.

**Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M. and Araki, T.** (2005). *TWIN SISTER OF FT (TSF)* acts as a floral pathway integrator redundantly with *FT*. *Plant and Cell Physiology*, **46**, 1175-1189.

**Yaniv, Z.** (1994) Preliminary report on major activities initiated within the framework of Network activities - Activities conducted in Israel. Rome, Italy:

**Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y. and Sasaki, T.** (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell*, **12**, 2473-2483.

**Yoo, S. D., Cho, Y. H. and Sheen, J.** (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols*, **2**, 1565-1572.

**Yoo, S. J., Chung, K. S., Jung, S. H., Yoo, S. Y., Lee, J. S. and Ahn, J. H.** (2010). BROTHER OF FT AND TFL1 (BFT) has TFL1-like activity and functions redundantly with TFL1 in inflorescence meristem development in Arabidopsis. *Plant Journal*, **63**, 241-253.

**Zhang, B., Wang, L., Zeng, L., Zhang, C. and Ma, H.** (2015). Arabidopsis TOE proteins convey a photoperiodic signal to antagonize *CONSTANS* and regulate flowering time. *Genes & Development*, **29**, 975-987.

**Zhang, J., Liu, G., Guo, C., He, Y., Li, Z., Ning, G., Shi, X. and Bao, M.** (2011). The *FLOWERING LOCUS T* orthologous gene of *Platanus acerifolia* is expressed as alternatively spliced forms with distinct spatial and temporal patterns. *Plant Biology*, **13**, 809-820.

**Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., Bogdanove, A. J. and Voytas, D. F. (2013).** Transcription Activator-Like Effector Nucleases enable efficient plant genome engineering. *Plant Physiology*, **161**, 20-27.

## **Appendix A.**

### **A.1. Primers used to isolate gene sequences**

#### **A.1.1. CONSTANS**

##### **A.1.1.1. Degenerate primers used for CDS isolation**

CO\_4\_F - ATGTTSAACAAGAGAGTAACRACAT

CO\_4\_R - TAGCAAGTGGGTTTGCAGAATGAA

CO\_5\_R - TGTATCTCAGGACCCTRGCTTCTC

##### **A.1.1.2. Primers used for RACE PCR**

GSP\_5'\_CO\_1\_R (RACE) - AGGGAACAACCATGAAGCCGTCTCTT

GSP\_5'\_CO\_2\_R (RACE) - CTGATGTGTGTTGAGTGGAGGCCAAG

GSP\_5'\_nest\_CO\_1\_R (RACE) - ACTCACAGACCGGGACACGTTTATGG

GSP\_3'\_CO\_1\_F (RACE) - AGATGTTCACCTCCGCAAACCTCGCTTG

GSP\_3'\_CO\_2\_F (RACE) - CTTGGCCTCCACTCAACACACATCAG

GSP\_3'\_nest\_CO\_1\_F (RACE) - CATGGACTACAAGTTCCCCGGTCAAT

##### **A.1.1.3. Primers used for full length CDS and gDNA isolation**

DtCO\_full\_2\_F - ATGTTGAAACAAGAGACTAACAACATG

DtCO\_full\_1\_R - CGTTATTTTTGACCGTAGGATGA

##### **A.1.1.4. GATEWAY primers**

DtCO\_full\_att\_2\_F -

GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTTGAAACAAGAGACTAA  
CAACATG

DtCO\_full\_att\_1\_R -

GGGGACCACTTTGTACAAGAAAGCTGGGTCGTTATTTTTGACCGTAGGAT  
GA

##### **A.1.1.5. Primers used in transgene amplification PCR**

pB2GW7\_prom\_1\_F - CGGATTCCATTGCCAGCTATCTGT

pB2GW7\_term\_1\_R - GATTTTTGCGGACTCTAGCATGGCC

#### **A.1.2. GIGANTEA**

##### **A.1.2.1. Degenerate primers used for CDS isolation**

GI\_1\_F - AAGCGACTTTTAGACGATGTGCTG

GI\_1\_R - TCAAACCATAAGCAGCTTCCCATA

GI\_2\_F - AGCTGCTGAAGATTATGCATCTGG

GI\_2\_R - ATACAGCTGAACCTCACAGGCAAG

GI\_3\_F - ATGTCTGAGCTTCGTGCTATGGTC

GI\_3\_R - TGAGGAAATTCGCAAGATCAGAAG

GI\_4\_F - ATGCATTGTCTGGGTTGATGAGAT

GI\_4\_R - AGATCTGATGCACTTGCGAGAATC

Dt\_GI\_1R - TCCATTTTGTGCTCCCTCTTG

Dt\_GI\_2R - AGGTTGGAGTGGGGTAGCAACA

#### **A.1.2.2. Primers used for RACE PCR**

GSP\_5'\_GI\_R (RACE) - CAACCGCCCATCTTGGTGTGGAAGG

GSP\_5'\_GI\_nest\_R (RACE) - CTGATACCAAGGGGAGCAGCAAG

GSP\_3'\_GI\_F (RACE) - TCAGATGAAGGGTCAGGGAGTAGA

GSP\_3'\_GI\_nest\_F (RACE) - GCAAGGATGGAGACAGGTAGTTGATG

#### **A.1.2.3. Primers used for full length CDS isolation**

Dt\_GI\_full\_1\_F - GATGACTTCTCCTACTTCTTCTGAG

Dt\_GI\_full\_1\_R - TTATTGGGACGACATTGATATGGTAC

#### **A.1.2.4. GATEWAY primers**

DtGI\_full\_comp\_att\_1\_F -

GGGGACAAGTTTGTACAAAAAAGCAGGCTGATGACTTCTCCTACTTCTTC  
TGAG

DtGI\_full\_comp\_att\_1\_R -

GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTGGGACGACATTGATAT  
GGTAC

#### **A.1.2.5. Additional primers used in sequencing reactions**

DtGI\_seq\_1\_R – CCAAGAGGCGAGTTCTGAATTTCT

DtGI\_seq\_2\_R – CTTTGCTCTCTGTATCTCCATTTTG

DtGI\_seq\_3\_R – ATATGATCCCTAGAAGTTTCTGGTG

DtGI\_seq\_1\_F – ATTCTGGCAAGTGCATCTGATCTTC

DtGI\_seq\_2\_F – AGCTACTGAAGATTATGCATCTGG

DtGI\_seq\_4\_R – TAAGGAAATTCGCTAGATCAGAAG

Dt\_GI\_RT\_1\_F – GATCTTCTTCTGAGAGCAACTGA

Dt\_GI\_RT\_1\_R – CTAACACTGGTTGTATTGCTCTTG

### **A.1.3. FLOWERING LOCUS C**

#### **A.1.3.1. Degenerate primers used for CDS isolation**

FLC\_1\_F - TACAGCTTCTCCTCCGGCGATAAC

FLC\_1\_R - CAGCTTCGGCTCCCACAAGATTAT

FLC\_2\_F - TCAAGATCCTTGATCGATATGGGA

FLC\_2\_R - GCTCTAGTTACGGAGAGGGCAGTC

FLC\_3\_F - GTGGATAGCAAGCTTGTGGAATCA

FLC\_3\_R - TAGTGGGAGAGTTACCGGAAGAT

#### **A.1.3.2. Primers used for RACE PCR**

GSP\_5' \_FLC1-3\_R (RACE) – TCCAACCTGAACGAGGGAACCGACAC

GSP\_5' \_FLC2\_R (RACE) – CTGCCCATGAAAAAGGGACAGAAG

GSP\_5' \_FLC2\_nest\_R (RACE) – GGGACAGAAGAAGCGGTAAAAGACAAGG

GSP\_3' \_FLC1-3\_F (RACE) – CCCTCTCCGTAACAAGAGCTAGGAA

GSP\_3' \_FLC1-3\_nest\_F (RACE) – CAATCTTCCGGTAACCTCCCCTGAA

GSP\_3' \_FLC2\_F (RACE) – GCCCTTGTCTTTTACCGCTTCTTC

GSP\_3' \_FLC2\_nest\_F (RACE) – GCAGCAAGCTTGTGGGATCAAATGTC

GSP\_3' \_FLC1-3\_nest\_F (RACE) – GCTGGAAGAGGAGAACCATGTTTTGG

#### **A.1.3.3. Primers used for full length CDS isolation**

Dt\_FLC\_fulllength\_1F – ATGGGGAGGAAGAACTTGA

Dt\_FLC\_full length\_1R – CTAATTAAGCAGTGGGAGTGTTAC

#### **A.1.3.4. Primers used for full length gDNA isolation**

DtFLC\_5'\_UTR\_1\_F – TGGTAACCCAACCTCAAGATCAAA

DtFLC\_3'\_UTR\_1\_R – GGATGCTTAAGGAAGCTCGAAGTA

#### **A.1.3.5. GATEWAY primers**

DtFLC\_full\_comp\_att\_1\_F -

GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGGGAGGAAGAACTTGA

DtFLC\_full\_comp\_att\_1\_R -

GGGGACCACTTTGTACAAGAAAGCTGGGTCTAATTAAGCAGTGGGAGTG  
TTAC

#### **A.1.3.6. Additional primers used in sequencing reactions**

DtFLC\_intron1\_in1\_F – TTCTTGTTGTCTCCGCCTCT

DtFLC\_intron1\_in1\_R – TCCCATAGTTGCACACTGGT

DtFLC\_intron1\_in2\_F – TTTACCCACACCTAATGAACAC

DtFLC\_intron\_1\_R – CCATGTCCATAGAAAAACAGTAGCA

DtFLC\_intron\_6\_F – AGCCTTTAGGCTCCTAGAGAAAGC

DtFLC\_exon1\_F – CTTCTCCAAACGACGCAATGGTCTC

DtFLC\_exon2\_F – GATCGATATGGAAAGCAACATGATG

DtFLC\_exon2\_R – CTTTAAGATCATCATCATGTTGCTTTC

DtFLC\_exon3\_F – CGTCAGTCAAACCTCTGAACTGTG

DtFLC\_exon3\_R – AGTTCCAGTAGCTCGTGGTGTGAACCA

DtFLC\_exon4\_F – GCTTGAGGAGTCAAACGTCGATAATG

DtFLC\_exon4\_R – CATTAGTTCTGTCTTCCTAGCTCTT

DtFLC\_exon5\_F – AGAAAAGGAGAAATTGCTGGAAGAGGAG

#### **A.1.4. SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1**

##### **A.1.4.1. Degenerate primers used for CDS isolation**

SOC1\_1\_F – TGGTGAGGGGCAAACCTCAGATGAA

SOC1\_1\_R – TGCAGCTAGAGCTTTCTCCTTTTGCTTG

SOC1\_2\_F – TGAATTCGCCAGCTCCAATATGCAA

SOC1\_2\_R – TGGGCTACTCTTTCATCACCTCTTCCA

##### **A.1.4.2. Primers used for RACE PCR**

GSP\_5'\_SOC1\_R (RACE) – GGTGCTGACTCGATCCTTGGTATG

GSP\_5'\_SOC1\_nest\_R (RACE) – GCACTGAGAGCTCAAAGGCTTTCTT

GSP\_3'\_SOC1\_F (RACE) – CGGAAACTCTTGGGAGAAGGCATAG

GSP\_3'\_SOC1\_nest\_F (RACE) – GAGCAGCTCAAGCAAAAGGAGAAAG

#### **A.1.4.3. Primers used for full length CDS isolation**

Dt\_SOC1\_full\_3\_F - GAGATGGTGAGGGGAAAACTCAGAT

Dt\_SOC1\_fulllength\_1R - GTTCACTTTCTTGAAGAACAAGGTA

#### **A.1.4.4. Primers used for full length gDNA isolation**

Dt\_SOC1\_5'UTR\_1\_F – ATA ACTACAAGGAGAAGGGATCCT

Dt\_SOC1\_3'UTR\_1\_R – GAGGCAGCTCAGTTTTAAAGAA

#### **A.1.4.5. GATEWAY primers**

Dt\_SOC1\_full\_att\_3\_F -

GGGGACAAGTTTGTACAAAAAAGCAGGCTGAGATGGTGAGGGGAAAAA  
CTCAGAT

Dt\_SOC1\_full\_att\_1\_R –

GGGGACCACTTTGTACAAGAAAGCTGGGTGTTCACTTTCTTGAAGAACA  
GGTA

#### **A.1.4.6. Primers used in transgene amplification PCR**

pB2GW7\_prom\_1\_F – CGGATTCCATTGCCAGCTATCTGT

Dt\_SOC1\_fulllength\_1R – GTTCACTTTCTTGAAGAACAAGGTA

#### **A.1.4.7. Additional primers used in sequencing reactions**

DtSOC1\_seq\_1\_F – TGATCACACACCCGTACACC

Dt\_SOC1\_UTR\_1\_R – TCCTTGGTATGCCTCAGATAACGA

### **A.1.5. FLOWERING LOCUS T**

#### **A.1.5.1. Degenerate primers used for CDS isolation**

FT\_1\_F – GGCCAAAGAGAGGTGACTAATGG

FT\_1\_R – TAACACACAATCTCATTGCCAAAGG

FT\_2\_F – TGGATCCAGATGTTCCAAGTCCTA

FT\_2\_R – CACGGGAAGGCCGAGATTGTA

#### **A.1.5.2. Primers used for RACE PCR**

GSP\_5' \_FT\_R2 (RACE) – GGAGGTGAGGGTTGCTAGGACTTGG

GSP\_5' \_FT\_nest\_R2 (RACE) – TGAACCTGAGAAGGCCTTAGATCC

GSP\_3' \_FT\_F2 (RACE) – CCTTTGGCAATGAGATTGTGTGTTA

GSP\_3' \_FT\_nest\_F2 (RACE) – TGCTGAGATCTACAATCTCGGCCTTC

### **A.1.5.3. Primers used for full length CDS isolation**

Dt\_FT\_fulllength\_2F - CATGTCTATAAATATAAGAGACCCTC

Dt\_FT\_fulllength\_1R - TCTAAAGTCTTCTTCCTCCGCAG

### **A.1.5.4. GATEWAY primers**

Dt FT full 2 att F –

GGGGACAAGTTTGTACAAAAAAGCAGGCTCATGTCTTATAATATAAGAG  
ACCCTC

Dt\_FT\_full\_2\_att\_R –

GGGGACCACTTTGTACAAGAAAGCTGGGTTCTAAAGTTCTTCCTCCGC  
AG

### **A.1.5.5. Primers used in transgene amplification PCR**

pB2GW7\_prom\_1\_F - CGGATTCCATTGCCAGCTATCTGT

Dt\_FT\_fulllength\_1R - TCTAAAGTCTTCTTCCTCCGCAG

## **A.1.6. TWIN SISTER OF FT**

### **A.1.6.1. Degenerate primers used for CDS isolation**

FT\_1\_F – GGCCAAAGAGAGGTGACTAATGG

FT\_1\_R – TAACACACAATCTCATTGCCAAAGG

FT\_2\_F – TGGATCCAGATGTTCCAAGTCCTA

FT\_2\_R – CACGGGAAGGCCGAGATTGTA

### **A.1.6.2. Primers used for RACE PCR**

GSP\_5'\_FT\_3\_R (RACE) – TTTTCCAAAAGCTGTTCCAGTTG

GSP\_5'\_FT\_nest\_R (RACE) – GAGATATTCTCGGAGGTGAGGGTTG

GSP\_3'\_FT\_F (RACE) – ATCCCTGCTACAACCTGGAACAACC

GSP\_3'\_FT\_nest\_F (RACE) – GTTTCGACAGCTTGGGAGGCAAA

GSP\_3'\_FT\_F2 (RACE) – CCTTTGGCAATGAGATTGTGTGTTA

GSP\_3'\_FT\_nest\_F2 (RACE) – TGCTGAGATCTACAATCTCGGCCTTC

### **A.1.6.3. Primers used for full length CDS isolation**

DtTSF\_fulllength\_1F – CATGTCTTTAAGTCAGAGAGATCCTC

DtTSF\_fulllength\_1R – TCTATGTTCTTCTTCCTCCGCAG

DtTSF\_5'UTR\_1\_F – GTTGAAAGTTGAGATCAAATATCATG

DtTSF\_3'UTR\_1\_R – CTTAAGAGGAACGAGGGATGGA

#### **A.1.6.4. GATEWAY primers**

Dt\_TSF\_full\_comp\_att\_1\_F –

GGGGACAAGTTTGTACAAAAAAGCAGGCTCATGTCTTTAAGTCAGAGAG  
ATCCTC

Dt\_TSF\_full\_comp\_att\_1\_R –

GGGGACCACTTTGTACAAGAAAGCTGGGTTCTAAAGTCTTCTCCTCCGC  
AG

#### **A.1.6.5. Primers used in transgene amplification PCR**

pB2GW7\_prom\_1\_F - CGGATTCCATTGCCAGCTATCTGT

DtTSF\_fulllength\_1R - TCTATGTTCTTCTCCTCCGCAG

## **A.2. Other Primers**

#### **A.2.1.1. RLM-RACE kit primers**

GSP 5' RACE primer F – CGACTGGAGCACGAGGACACTGA

GSP 5' RACE nested primer F – GGACACTGACATGGACTGAAGGAGTA

GSP 3' RACE primer R – GCTGTCAACGATACGCTACGTAACG

GSP 3' RACE nested primer R – CGCTACGTAACGGCATGACAGTG

#### **A.2.1.2. Primers for colony screening and clone sequencing**

M13 F – TGTAACAACGACGGCCAGT

M13 R – GGAAACAGCTATGACCATG

#### **A.2.1.3. Primers for rocket real time PCR**

DtTIP41\_RT\_1\_F – CCATACTGTGGTAGCGGAGTTGTT

DtTIP41\_RT\_1\_R – AGACTAGCTTCCCCCTCAGGACTT

DtCACS\_RT\_1\_F – CCTAGTGGTCGGAGAAAAGAGGAA

DtCACS\_RT\_1\_R – GTGCCGTCAATCTCAGAAGTCAGT

Dt $\alpha$ -tubulin\_RT\_1\_F – GTTCCCAAAGACGTTAATGCTG

Dt $\alpha$ -tubulin\_RT\_1\_R – TTGGAGGTTGGTAGTTGATTCC

DtFLC\_RT\_1\_F – CGTAACAAGAGCTAGGAAGACAGA

DtFLC\_RT\_1\_R – CGCAAGATTATTCTTCTCCATC

DtFT\_RT\_1\_F – GTGTGTTACGAAAATCCAAGTCC  
DtFT\_RT\_1\_R – GATTGTAGATCTCAGCAAACCTCG  
DtSOC1\_RT\_1\_F – AAAGACTCAAGTGTTTAAGGAT  
DtSOC1\_RT\_1\_R – ACTTTCTTGGTTCTTATTTGAC  
DtCO\_RT\_1\_F – CGGTAACCTGGAACTCTTACACC  
DtCO\_RT\_1\_R – GTTTTTAGGGAACAACCATGAAG  
DtGL\_RT\_1\_F – GATCTTCTTCTGAGAGCAAAGTGA  
DtGL\_RT\_1\_R – CTAACACTGGTTGTATTGCTCTTG  
DtTSF\_RT\_1\_F – CCGAGTTGCATTGATGTTGTTCC  
DtTSF\_RT\_1\_R – CCATTCTCTCTCTGACAGTTGAAG

**A.2.1.1. Primers used for Arabidopsis complementation of *DtFT* T<sub>3</sub> real time PCR**

At\_ACTIN2\_RT\_1\_F – TGTCGCCATCCAAGCTGTTCT  
At\_ACTIN2\_RT\_1\_R – GTGAGACACACCATCACCAGAAT  
At\_TIP41\_RT\_1\_F – ATGGTGTGCTTATGAGATTGAGAG  
At\_TIP41\_RT\_1\_R – ATACCCTTTTCGCAGATAGAGACTG  
AtTUBBY\_RT\_1\_F – TGGCAAGATGAGCACAAAAG  
AtTUBBY\_RT\_1\_R – AGACCTCGGGGAGCTATG  
Dt\_FT\_RT\_1\_F – GTGTGTTACGAAAATCCAAGTCC  
Dt\_FT\_RT\_1\_R – GATTGTAGATCTCAGCAAACCTCG

**A.2.1.2. Primers used for Arabidopsis complementation of *DtFLC* T<sub>2</sub> real time PCR**

At\_ACTIN2\_RT\_1\_F – TGTCGCCATCCAAGCTGTTCT  
At\_ACTIN2\_RT\_1\_R – GTGAGACACACCATCACCAGAAT  
At\_TIP41\_RT\_1\_F – ATGGTGTGCTTATGAGATTGAGAG  
At\_TIP41\_RT\_1\_R – ATACCCTTTTCGCAGATAGAGACTG  
AtTUBBY\_RT\_1\_F – TGGCAAGATGAGCACAAAAG  
AtTUBBY\_RT\_1\_R – AGACCTCGGGGAGCTATG  
DtFLC\_RT\_2\_F – GGTTCCCTCGTTCAGTTGGAA  
DtFLC\_RT\_2\_R – CTCCATCTGGCTAGCCAAAAC

## Appendix B.

### B.1.1. Gene and protein sequences of isolated flowering time genes in rocket

#### B.1.1.1. CDS for DtCO

ATGTTGAAACAAGAGACTAACAACATGAGTAATAGAGAGGCACGAGCCTGTGACACA  
TGCCGTTCaACCATATGCACCGTGTACTGCCATGCTGACTCAGCCTACTTATGCAAT  
AGCTGTGATGCTCAAGTCCACTCAGCCAACCGCGTTGCTTCCCGCCATAAACGTGTC  
CCGGTCTGTGAGTCATGTGAGCGTGCCCTGCTGCTTTTTATGTGCGAAGCAGATGAC  
GTGTCTCTATGCACAGCCTGTGATGAAGATGTTCACTCCGCAAACCTCGCTTGCAAAA  
CGCCATCATCGAGTTCCCGTTGTGCCGGTAACTGGAACTCTTACACCCCCTTGGCC  
TCCACTCAACACACATCAGTGACCGAGCCAGAGAAGAGAGTGGTGTTAGAGGATGAT  
AAAGAGACGGCTTCATGGTTGTTCCCTAAAAACAGTGACGATCACAACAACAACAAC  
AACCAGAACAATGAGTTGTTGTTTAATGATGACTATCTAGACCTTGCTGATTACAAC  
TCCAGCATGGACTACAAGTTCCCGGTCAATACAATCAACATCAACATAAACAAGAC  
TGCACCGTACCACAGACAACTACGGTGGTGATAGAGTTGTTCCACTCAA AATTGAA  
GAAACAAGAGGAAACGTGCACCACAAGCAACATAATATCACATATGGCTCCTCAGAG  
TGTCACTACAAC TACAATGGTTCCATAAACATAACGCATACAATCCATCAGTGAA  
ACTGACTTTGTCCCGGAGCCGACACGTGACACAACAGTTTCACACCAAAGACTCCC  
AAAGGGAACAGAGAAGCCAGAGTCCTGAGATACAGAGAGAAGAAGAAGAGAAGAAAG  
TTTGAGAAGACAATAAGGTATGCTTCAAGGAAGGCATATGCAGAGAGAAGACCAAGG  
ATCAATGGACGATTTGCAAAGATGAGTGAAACTGAAGCAGAGGAGCAAGAGTTCAAC  
AGAATGCTAATGTACTATGACACTACAGGATATGGCATTGTTCCCTTCATCCTACGGT  
CAAAAATAA

#### B.1.1.2. gDNA for DtCO (black – non coding sequence, red – coding sequence)

ATCATCTCACTTACTACATATCTGAAAGTTAACATGTTGAAACAAGAGACTAACAAC  
ATGAGTAATAGAGAGGCACGAGCCTGTGACACATGCCGTTCAACCATATGCACCGTG  
TACTGCCATGCTGACTCAGCCTACTTATGCAATAGCTGTGATGCTCAAGTCCACTCA  
GCCAACCGCGTTGCTTCCCGCCATAAACGTGTCCCGTCTGTGAGTCATGTGAGCGT  
GCCCTGCTGCTTTTTATGTGCGAAGCAGATGACGTGTCTCTATGCACAGCCTGTGAT  
GAAGATGTTCACTCCGCAAACCTCGCTTGCAAAACGCCATCATCGAGTTCCCGTTGTG  
CCGGTAACTGGAACTCTTACACCCCCTTGGCCTCCACTCAACACACATCAGTGACC  
GAGCCAGAGAAGAGAGTGGTGTTAGAGGATGATAAAGAGACGGCTTCATGGTTGTTT  
CCTAAAAACAGTGACGATCACAACAACAACAACAACCAGAACAATGAGTTGTTGTTT

AATGATGACTATCTAGACCTTGCTGATTACAACCTCCAGCATGGACTACAAGTTCCCC  
GGTCAATACAATCAACATCAACATAAACAAGACTGCACCGTACCACAGACAAACTAC  
GGTGGTGATAGAGTTGTTCCACTCAA AATTGAAGAAACAAGAGGAAACGTGCACCAC  
AAGCAACATAATATCACATATGGCTCCTCAGAGTGTCACTACAACCTACAATGGTTCC  
ATAAACCTAACGTAAGACTTTTTTTTTTATATATTTTCTCCTCTTTTCATTTTACAC  
GCAAATCACATATTAATTACTAACAATTGATAGGTGGTACATAGA AACTTTTTATAAT  
AATTA AAAAAGGTTGCTTCTCTTTAGTCCTCTTACCAAGGGGATTGATCAGAATTGT  
CTTTGTTTTTTTTTTTAGGCATACAATCCATCAGTGGAACTGACTTTGTCCCGGAGC  
CGACACGTGACACAACAGTTTCACACCAAAGACTCCCAAAGGGAACAGAGAAGCCA  
GAGTCTGAGATACAGAGAGAAGAAGAAGAGAAGAAAGTTTGAGAAGACAATAAGGT  
ATGCTTCAAGGAAGGCATATGCAGAGAGAAGACCAAGGATCAATGGACGATTTGCAA  
AGATGAGTGAACTGAAGCAGAGGAGCAAGAGTTCAACAGAATGCTAATGTACTATG  
ACACTACAGGATATGGCATTGTTCTTTCATCCTACGGTCAAAAATAACGAGTACTAG  
ATTGCCAAAAACCTGTAA AATACTGTTTAGTTTGAGATTATGAGGTCTGGTGA AAT  
GATTACTTATTCCGCTAT

### **B.1.1.3. Protein sequence of DtCO**

MLKQETNMSNREARACDTCRSTICTVYCHADSAYLCNSCDAQVHSANRVASRHKRV  
PVCESCERAPAAFMC EADDVSLCTACDEDVHSANSLAKRHHRVPVVPVTGNSYTPLA  
STQHTSVTEPEKRVVLEDDKETASWLFPKNSDDHNNNNNQNNE LLFNDDYLDLADYN  
SSMDYKFPGQYNQH QHKQDCTVPQTNYGGDRVVPLKIEETRGNVHHKQHNITYGSSE  
CHYNYNGSINHNAYNPSVETDFVPEPTRD TTVSHQKTPKGNREARVLR YREKKKRRK  
FEKTIRYASRKAYAERRPRINGRFAKMSETEAEEQEFNRMLMYDYDTTGYGIVPSSYG  
QK

### **B.1.1.4. CDS of DtFLC**

ATGGGGAGGAAGAACTTGAAATCAAGCTAATCGAGAACAAAAGTAGCCGACAAGTT  
ACCTTCTCAAACGACGCAATGGTCTCATCGAGAAAGCTCGTCAGCTGTCCGTTCTT  
TGTGACGCATCCGTTGCTCTTCTTGTGTCTCCGCCTCTGGAAAGCTCTACAGCTCC  
TCCTCCGGTGATAACCTGGTCAAGATCTGTGATCGATATGGAAAGCAACATGATGAT  
GATCTTAAAGCCTTGGATCGTCAGTCAAACCTCTGAACTGTGGTTCACACCACGAG  
CTACTGGA AACTTGTGGAAAGCAAGCTTGAGGAGTCAAACGTGATAATGTAAGTGT  
GGTTCCTCGTT CAGTTGGAAGAGCACCTTGAAAACGCCCTCTCCGTAACAAGAGCT  
AGGAAGACAGAACTAATGTTGAAGCTTGTGGAGAACCTTAAAGAAAAGGAGAAATTG  
CTGGAAGAGGAGA ACCATGTTTTGGCTAGCCAGATGGAGAAGAATAATCTTGCGAGA

GCTGAAGCTGATGATATGGAGGTGTCACCTGGACAAATCTCCGACATCAATCTTCGG  
GTAACACTCCCCTGCTTAATTAG

**B.1.1.5. gDNA of DtFLC (black – non coding sequence, red – coding sequence)**

GAANCAAAATAANACGCTCAATATCTCCGGNGACTTGGTAACCCAACCTCAAGATCA  
AATTAGGGCACAAAGCCCTGTCGGAGACACCAAGCCATGGGGAGGAAGAACTTGAA  
ATCAAGCTAATCGAGAACAAAAGTAGCCGACAAGTTACCTTCTCCAACGACGCAAT  
GGTCTCATCGAGAAAGCTCGTCAGCTGTCCGTTCTTTGTGACGCATCCGTTGCTCTT  
CTTGTGTCTCCGCCTCTGGAAAGCTCTACAGCTCCTCCTCCGGTGATAAGTACGAC  
TTCTCCTCTTCTGGGTTTTATATTTTTCTCTTGTTAATTACTAACCTTTTCAAGAAA  
AGTATAACTATTATTTTTTTTTTTTTAAATATTTCTGCGAATCTAGTGKGGTTTTAT  
CAGTTTGTGTGTGCTACTGTTTTCTATGGACATGGGARATTCCAAATCAAATGAA  
GAACAATGTTGTGATGAGCTAAAGCTTCTTACTTTTGTCTTTTCTTTTCTCTCTCTC  
TTCTTCTTCTTCTTCTCACAATTTCTGCATGGATTTTTTTTTTATTGGTAAAAAAT  
TGCATGTCCCTCAARATCTTTTGTACACGTTGTGATGCATAACTGATGTTGCTTAGT  
GAAGTTTATATCCATCTTTGATTGTTAGTTCATCAATTGTTGTTTAGCTGTAGGAAC  
TCCTTTTGTGTTGARAAGGGTAAATGATTCATGTTTCTTTATGTTGTTAGGGTAATTA  
AATCCATTGGATCTCTCAAATCTGTATGCAGTGAAATGCACCTTGGGGAGATCTATA  
GAATGGCTGTAAGCTTGATACTGAATAGGTTCCCTAACTGATTTTAAATGKGCATAGT  
AGTTTTGAATTTTGGTAGCTTCTAAAAATACAGATTCACAATGAGTTACTTAGTTGA  
TCCAGAGTATCGATAGCGGTTCTTTTGTATCGCAAGCGTTGTGTGCGTATATAAGTCT  
TTACCCACACCTAATGAACACTAATTCGTCTATTTTTTAGACTAATTAATTTTTATA  
TTTCTCTTGATTTTTCGTACAATGCAATGCAAATCCATGTAGATAAAATTAGTGGAA  
TTGATGGTATATTAATAGATCATCAAGTCTGCGCTTATAGGKACACCCAGGTGTGCA  
ACTAKGGACTAAAATATTTTAGGTCTAAATGTTTTCTARAAGTTTTCYCTCCCTTTC  
ACTGATTTGTTGATATAAGATTATCTCAAGCTTTTAGCTTCGCACCCACATATTTAC  
GGCACATGCATATGCTACATTAATATTGTGTAATTGGATTATAGCTTTGTTCAATTTG  
AATTGACAACCTACAACATTGTAGCTTTTTATTATGGTAATGTACAATCACACAACCT  
TGCATCTGTTTGTGTTTCTTGTGTGGAAATTGACAATAACCAAACGTGCTTCCTTTC  
GTGCCTCTTGGAAACCAACAACCTCACTGAAACCATTTTCTTCATGGGAAAATTCCA  
ACCCCTTATCGAAATTTGAATTTGTTGTAGGGTTTACTCAACCCTATAATATTAACC  
AACTGGTTGTAGTGTTTTGGCAATGTTAGTGAAGATCGCAGGTCATGTCTCACTTTA  
GTTAATACGTGTTCACTCTATGAGTGTTAGGTCCAGATGTGGAATAGCTAGTCGAGA  
TTAGTTATATGTGAGTAAAAGAAGAAAGGCTTTCAGAACCTGGAATTTGACCTTTAG  
TTACTCTATAACGCATTTGTGCAAAAAGCTTAACTTCTAACACAAAACCTGTTGAATT

AATGGAAC TGT TTTTCAAAT TTTGCTTGGACTTCTATAAACTCATAAATAAGAAAAT  
TTTGTCCGTCTTCAGCTGTTGCTTCTATGGATTACCTTTACGCCCTAATAGTTCCGT  
TTGCATTTACGTTTACGATTGCGTTTCCGTGTTCTAGTTTTCTAATAACTAGAATGT  
CTTAAGCACAAACGTGTGGCAAGGAAAGAAACAATTGGGGCCTGTGATTAGGGTTCC  
TGTGTACCTTACTTTTTACTGCTTAAAAATCATAAGAGTGAAAGTTACAAAATAA  
GAAACTAGTATTTACATAAACTTAAATACTAGTTTTCTGCAATGATAATGAAAATA  
TATATGTGAAAATCAAATTTTGTCTCTAATTGTCTCTTCCGTGCCCTATACATGYTT  
TATAGTTTCCAGTGCGGCTTTCAAGGGTGGCGGCATSTTTGMTCCTAACATATTTTG  
ATWTTMTYTTGTCTTCTCTCCAGCCTGGTCAAGATCTGTGATCGATATGGAAAGCAA  
CATGATGATGATCTTAAAGCCTTGGTAATACAAATATCTTTTATTTGATGAAATTTG  
TTAGACTAAATTTAATTATGGCGTAGAAGTTTGTTCATTTTGATCAGTAGCAATTC  
ATGTAGCTTTGAGGTTTGATTTTTTTGATTGAGGGGAGGGGTATAGAAGGCATTAGC  
ACGCACGTACTTAGCGTTACACATATAATTACTATTTTGGAGCTGTAGGTATCAAAG  
TTTGT TTTTCAAGATCATGAAAGTTTGATGGTCCTTTAGGGCTAGTTGTTTGATAGGA  
TTGGGTAAACACATGTTTGGCACCTCAAATTTGTGGAGATCTAGTTTTAGGAATTGC  
GGTATTAATCAGTTTATTATTTATTACTTTCCTTAATGATCCCTGTATATGAGATGA  
TGGATTCTCATGGATAACTAATTTTCAAAGCTTATATCACTTCTCCGTGGACAGGAT  
CGTCAGTCAAACCTCTGAACTGTGGTTCACACCACGAGCTACTGGAACCTGTGGAA  
AGTTAGTGTTAACTAAAACCTTTTTCTCTCCTTTGATCACAAAGGAATTAGGGTTTC  
TTGTCAATCTATGAATATTTCCAGCAAGCTTGAGGAGTCAAACGTCGATAATGTAAG  
TGTCGGTTCCTCGTTCAGTTGGAAGAGCACCTTGAAAACGCCCTCTCCGTAAACAAG  
AGCTAGGAAGGTACGTTGCTTCGGTGATGTCTTCTTTTTCTTACTTTGGTTTTAAA  
GATCAACCGTGCCTTATATTTATTACTTTTGTTCAGACAGAACTAATGTTGAAGCT  
TGTGGAGAACCTTAAAGAAAAGTTAGATATATGCTTCCAAGATTATAGCACTTAAG  
TTATATAATACATATTGTAGTCGTGTTGTGTTAGTTTATCCAAGTCTGCTATCATAA  
GGTCGCGAACCTATCTCAAACCTATTGCCAATGGATCTTTAGTCTTTTCAGGATCAT  
TAGAATGTTTTGAATGAAAAAATCTGATGTTAGGTCTATAAACTTTATCTATGAGTA  
AAATGATGTGTTGATCTCTAAACAGGAGAAATTGCTGGAAGAGGAGAACCATGTTTT  
GGCTAGCCAGGTAAC TAAAGTTTTTTTTTTGTAACGGTTTAAGTTTGTACCCATTGTCA  
CAGTTCAAGTTTATAATGCTACTTTTACCTAAGGCTAATTATATACTAGAAAAGGG  
AAGATTACATGGTCTGGGGAACCTAAACCAAAAAAAGGGAACCTAAACCAAAATGA  
AGCCTTTAGGCTCCTAGAGAAAGCTATTTTGCAGGCTTATGCTTGCAAAC T TAAAAT  
TAAAAGCTCAAAGTCAGAAACAGAAATTTGTGAAACAAGATTCATATTGGTTATCAG  
TGTAATTTGTGGTAGGAAAATAATATTTAATTAATATAGGATGACATAGAGAGAGA

AGTGAATAGTGTTTTAATTATTTCTAAAGAGATGAAAACCTATGTTTGACCTAATATA  
TATTTATGTTACAGATGGAGAAGAATAATCTTGCGAGAGCTGAAGCTGATGATATGG  
AGGTGTCACCTGGACAAATCTCCGACATCAATCTTCGGGTAACACTCCCCTGCTTA  
ATTAGTCAACCTTTAATCGGCGAATGAATAAAAAGCCAAAACATATAAGTAAAGCAAG  
ATGATGTGTAATTATCCCCTTGTAAGGGTGTACGTTGTATAATCATAATACTCTTT  
CCTGCCGAGAGGCTTCGGGTGTAAGGATTTGAGATTTATGTAAAATAGAAAATCT  
ATGTAAGATACTTTGAACTTTGTACCTTTGTACCTTTGTAGAGAGAGATAGAGTTA  
CTTCGAGCTTCTTTAAAGCAT

#### **B.1.1.6. Protein of DtFLC**

MGRKKLEIKLIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSS  
SSGDNLVKICDRYGKQHDDDLKALDRQSKPLNCGSHHELLELVESKLEESNVDNVS  
GSLVQLEEHLENALSVTRARKTELMLKLVENLKEKEKLLLEENHVLASQMEKNNLAR  
AEADDMEVSPGQISDINLRVTLPLLN

#### **B.1.1.7. CDS for DtFT**

ATGTCTATAAATATAAGAGACCCTCTTATAGTAAGCAGAGTTGTTGGAGACGTTCTT  
GATCCGTTTAATAGATCAATCACTCTAAAGGTTACTTATGGCCAAAGAGAGGTGACT  
AATGGCTTGATCTAAGGCCTTCTCAGGTTCAAACAAGCCAAGAGTTGAGATTGGT  
GGAGAAGACCTCAGGAACCTTCTATACTTTGGTTATGGTGGATCCAGATGTTCCAAGT  
CCTAGCAACCCTCACCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCT  
ACAACCTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCAAGTCCCCT  
GCAGGAATTCATCGTGTCTGTTTATATTGTTTCGACAGCTTGGGAGGCAAACAGTG  
TATGCACCAGGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT  
CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAGAGGGAGAGTGGCTGCGGA  
GGAAGAAGACTTTAG

#### **B.1.1.8. Protein sequence of DtFT**

MSINIRDPLIVSRVVDVLDVDFNRSITLKVTYGQREVTNGLDLRPSQVQNKPRVEIG  
GEDLRNFYTLVMVDPDVPSPSNPHLREYLHWLVTDIPATTGTTFGNEIVCYENPSPT  
AGIHRVVFILFRQLGRQTVYAPGWRQNFNTREFAEINYNLGLPVAAVFYNCQRESGCG  
GRRL

#### **B.1.1.9. CDS for DtGI**

ATGACTTCTCCTACTTCTTCTGAGAAGTGGACCGATGGTCTTCAGTTCTCTTCCTTG  
TTATGGTCTCCTCCACGTGACCCTCAACAACATAAGGATCAAGTCGTTGCTTATGTT  
GAATACTTTGGTCAGTTCACATCAGAGCAATTCCTGATGATATTGCTGAGTTGGTC

CGGAATCAGTATCCATCAACTGAGAAGCGGCTTTTGGATGATGTGTTGGCTATGTTC  
GTCCTCCATCATCCTGAGCATGGTCATGCGGTCATCCTTCCAATCATTTTCATGTCTC  
ATCGATGGCTCTTTAGTGTACAGCAAGGAAGCTCATCCTTTTGCCTCTTTCATATCT  
TTAGTTTCCCAAATAGTGAGAATGACTATTCAGAGCAATGGGCTTTGGCGTGCGGA  
GAAATTCTCCGCATTTTGACCCATTACAATCGTCCCATATACAAGAGGGAGCAACAA  
AATGGAGATACAGAGAGCAAAGCTTCTACTAGTGGCAGTGGGTCTCCTACGCTTTCA  
GAGACTAAGGTTGCATCACCAGGACAGAATGAAAGGAAACCGTTAAGGCCTTTGTCT  
CCTTGGATCAGTGATATACTACTTGTGCTGCCCTTGGTATCAGAAGTGACTACTTT  
CGTTGGTGTAGTGGTGTATGGGTAAATATGCTGCTGGAGAGCTCAAGCCACCTACC  
ATTGCTTCTCGAGGATCTGGTAAACATCCTCAATATATGCCCTTCCACACCAAGATGG  
GCGGTTGCTAATGGAGCTGGTGTACACTGAGTGTGTGTGATGATGAAGTTGCTCGG  
TATGAGACTGCTACGCTAACAGCTGTTGCTGTCCCTGCACTTCTTCTTCCCTCCCCA  
ACGACATCTTTAGATGAGCATTAGTTGCTGGCCTTCCAGCTCTTGAGCCATATGCA  
CGTTTGTTCATAGATATTATGCTATTGCAACTCCAAGTGCTACTCAGAGACTTCTT  
CTTGGACTCTTAGAAGCACCACCGTCTTGGGCTCCAGATGCACTAGATGCTGCTGTA  
CAGCTTGTGGAACCTCTTTCGAGCTACTGAAGATTATGCATCTGGTGTAAAGGCTTCCC  
AGGAATTGGATGCATTTGCACTTCTTGCCTGCAATAGGAATTGCTATGTCTATGAGG  
GCAGGTGTGGCTGCTGATGCTGCAGCTGCTTTACTTTTCCGCATACTCTCACAACCG  
GCACTGCTTTTTTCTCCACTAAGCCAAGCTGAAGGAGTAGAAATTCAGAACTCGCCT  
CTTGGTGGCTACAGTTCAAATCATAGAACGCAGGCAGAAGTTTCAGCAGCAGAAGCA  
ACCATTGAAGCCACTGCTCAAGGAATTGCCTCAATGCTTTGTGCTCATGGTCCTGAA  
GTTGAGTGGAGAATCTGCACTATATGGGAAGCTGCTTATGGATTGATCCCTTTAAAC  
TCCTCAGCGGTTGATCTTCCAGAAATCATTGTTGCTACCCCACTCCAACCTCCTATC  
TTGTCATGGAATCTGTACATTCCACTCCTCAAAGTACTTGAGTATCTTCCACGTGGG  
AGTCCTTCCGAGGCATGCTTGTGATGAAAATATTTGTTGCCACTGTGGAAACAATCCTC  
AGCAGAACTTTCCACCAGAACTTCTAGGGATCATATCAGGAAAGCTAGATCGAGT  
TTAACCACGAGATCAGCGACCAAAAACCTTGCTATGGCTGAGCTTCGTGCTATGGTC  
CATGCTCTTTTTTAGAATCATGCGCTGGTGTGGAGATAGCTTCCAGCTTGCTTTTT  
GTTGTGTTGACTGTATGTGTTAGCCATGAAGCACAGTCTAGTGGGAGCAAGAGACGG  
AGAAGCGAAGATGCTAGTAGCACTGCAGAGGAGAATCAACATGTGTCTGACAATCAA  
ACTAGTAACCGTAAAAGTAGGAGTGTCAAGGGACAAGGACCTGCGGCAGCATTGAT  
TCATACGTTCTCGCTGCTGTGTGTGCTCTTGCCTGTGAGGTTCCAGCTGTATCCTATG  
ATCTCTGGAGGGGGAAACTTTTCCAACCTCTGCAGTGGCTGCAACCATTACAAAGTCT  
GTAAAGATAAATGGTTCATCTAACGAGTACGGATCTGGGATTGACTCTGCAATTAAG

CATACACGCCGTATCTTGGCGATTCTTGAGGCACTCTTTTCATTGAAACCATCTTCT  
GTGGGGACTCCATGGAGTTACAGCTCAAGCGAGATAGTTGCTGCGGCCATGGTTGCA  
GCTCATATCTCAGAACTGTTTCAGACGTTCAAAGGCCTTGACGCATGCGTTGTCTGGG  
TTGATGAGGTGCAAATGGGACAAGGAGATTCATAAAAGAGCATCGTCGTTGTATAAC  
CTCATCGATGTTTCACAGCAAAGTTGTAGCATCCATTGTAGACAAAGCTGAACCCTTA  
GAAGCATAACCTTAAGAATGCACCTCTCCAGAAGGATGCGCTGACTTGTCTTAACTGG  
AAACAACAGAACACATCAGGGTTTGGTACAGCGGCGGTGATGTCCACGTCACGTAAT  
GAAATGACTCCAAGAGGAGGTAACAACCATAAGTGTGCTAGACATTCAGATGAAGGG  
TCAGGGAGTAGATCCTCCTCCTCAGAGAAGGGTATCAAAGATCTGCTGTTGGATGCT  
TCTGATCTAGCGAATTTCCCTTACAGCTGATAGGCTCGCAGGGTTTTACCGTGGTACA  
CAAGTGCTTTTTGAGGTCTATACTTGCAGGAGAAACCTGAGCTTTCTTCTCCGTTGTT  
TCGCTGTTGTGGCATAAACTGATCGCTTCTCCTGAAATCCAGCCCACCGTTGAGAGC  
ACCTCTGCTCAGCAAGGATGGAGACAGGTAGTTGATGCACTATGCAATGTCGTATCT  
GCAACGCCAGCAAAGCAGCTGCAGCCGTTGTTCTTCAGGCCGAGAGAGAGTTGCAG  
CCATGGATAGCGAAAGATGATGAAGAAGGTCAGAAAATGTGGAAAATAAACCAAAGG  
ATAGTGAAAGTGATGGTGGAACTCATGAGAAACCATGACAGGCCTGAGTCACTGGTG  
ATTCTSGCAAGTGCATCTGATCTTCTTCTGAGAGCAACTGATGGAATGCTTGTGGAC  
GGAGAAGCTTGTACATTACCTCAACTTGAGCTACTTGAAGCCACGGCAAGAGCAATA  
CAACCAGTGTTAGCTTGGGGACCATCTGGACTAGCAGTGGTTGATGGTTTATCCAAT  
CTATTGAAGTGTCGTCTACCGGCAACAATAAGGTGCCTGTCACACCCAAGCGCACAC  
GTACGTGCCTTGAGCACATCAGTACTACGAGACATCATGAACCAAAGCGGCATAACC  
ACCGCCACCACCAAGGCAACACCAAACCACCGCCAACAATAACAACAGAGAAGAAC  
GGAACGGACAGTCCGTCCTACCGGTTCTTCAACGCGGCTGCAATAGACTGGAAAGCA  
GACATACAAAAGTGTTTGAAGTGGGAAGCTCACAGCTTGCTGTCCACGACAATGCCA  
ACACAGTTTCTTGACACGGCGGCTCGGGAATTAGGATGTACCATATCAATGTCGTCC  
CAATAA

**B.1.1.10. gDNA of DtGI (black – non coding sequence, red – coding sequence)**

AGAACTCTATAACTCTTCAGCTTCGATTTAGTGTAGTATAGTAGATCAAGAACTGTT  
ACGACTTATTGGTTAGAAAAACAGGCTGTAACCATATTATCCCTAG**ATGACTTCTC**  
**CTACTTCTTCTGAGAAGTGGACCGATGGTCTTCAGTTCTCTTcCTTGTTATGGTCTC**  
**CTCCACGTGACCCTCAACAACATAAG**GTACTACTAACAATAATAAAGAAGCATCAGT  
TTTTTTTATAAGCTGTCTCATCCATGAACCGTCTTGATATGTGGTTACTGATTGTC  
TTTCTGCTTTAG**GATCAAGTCGTTGCTTATGTTGAATACTTTGGTCAGTTCACATCA**

GAGCAATTCCTGATGATATTGCTGAGGTACTTGTTGTGTCTTATGTTGTTTATTGT  
AATTTACTGAATAAAAGCTTCTCTCAGTAGTGATGATTTACTGAATTTATATAAATT  
ATCTCTGTATCAGTTGGTCCGGAATCAGTATCCATCAACTGAGAAGCGGCTTTTGG  
TGATGTGTTGGTATGTATGAGCTTCTATATCAAACACTTATGATCCATACTTCTTT  
GTTTTGAATGGTTTAATCTGATTCTCTGGCTTGCTGATGAAACAACAGCTATGTTCCG  
TCCTCCATCATCCTGAGCATGGTCATGCGGTCATCCTTCCAATCATTTTCATGTCTCA  
TCGATGGCTCTTTAGTGTACAGCAAGGAAGCTCATCCTTTTGCCTCTTTTCATATCTT  
TAGTTTCCCCAAATAGTGAGGTAAAAATAGTTTCTCACAAGTGTTAGTTTAAGACTT  
GTCTCTGTGGAAAGTGGTAACTCTTTTTTTTTTTGGTATGCTGACATAAATGTTTTTC  
ACGATCTTAACAGAAATGACTATTTCAGAGCAATGGGCTTTGGCGTGCGGAGAAATTC  
CCGCATTTTGACCCATTACAATCGTCCCATATACAAGAGGGAGCAACAAAATGGAGA  
TACAGAGAGCAAAGCTTCTACTAGTGGCAGTGGGTCTCCTACGCTTTCAGAGACTAA  
GGTTGCATCACCAGGACAGAATGAAAGGAAACCGTTAAGGCCTTTGTCTCCTTGAT  
CAGTGATATACTACTTGCTGCTCCCCTTGGTATCAGAAGTGACTACTTTCGTTGGTA  
AGTAGATTATGTTACTTGATATCTTGCTCTCAGGGACTGTGGATTAAAGTTCTAT  
AGTGTGCTTAAGCTTCAACTAAAAAGTTGTTGTATATATCTACAGGTGTAGTGGTG  
TTATGGGTAAATATGCTGCTGGAGAGCTCAAGCCACCTACCATTGGTGAGTTTCCAT  
CACCTGTGTGGCTATTAGAATGGGTTGTGAGTTACGTCTTAAACACTCACGCTTAG  
TACTTTCCTACTTTTGCAGCTTCTCGAGGATCTGGTAAACATCCTCAATATATGCCT  
TCCACACCAAGATGGGCGGTTGCTAATGGAGCTGGTGTACTACTGAGTGTGTTGTGAT  
GATGAAGTTGCTCGGTATGAGACTGCTACGCTAACAGCTGTTGCTGTCCCTGCACTT  
CTTCTTCTCCCCAACGACATCTTTAGATGAGCATTTAGTTGCTGGCCTTCCAGCT  
CTTGAGCCATATGCACGTTTGTTCATAGGTAGTGTCTTAAACACACCCTTTGGAT  
TGTGCTTACTGGATTTCATAGGCTGTTTAGAATTGCTTCTCACTTCTTGTCTAATA  
ACAGATATTATGCTATTGCAACTCCAAGTGCTACTCAGAGACTTCTTCTTGGACTCT  
TAGAAGCACCACCGTCTTGGGCTCCAGATGCACTAGATGCTGCTGTACAGCTTGTGG  
AACTCCTTCGAGCTACTGAAGATTATGCATCTGGTGAAGGTAAGCGTAATATGAA  
TCCTTTAaTCACTCTCaAGTCTCAACAAAGATGATTGTGTCCCTTGTTACACACTAT  
ATATGAGATTTGTTGGCATAACATTGTGTCATGGATTAGCTCTGAATACACTCTCTTT  
TCTTTATTTAGCTTCCAGGAATTGGATGCATTTGCACCTTCTTGCCTGCAATAGGAA  
TTGCTATGTCTATGAGGGCAGGTGTGGCTGCTGATGCTGCAGCTGCTTTACTTTTCC  
GCATACTCTACAACCGGCACTGCTTTTTCTCCACTAAGCCAAGCTGAAGGAGTAG  
AAATTCAGAACTCGCTCTTGGTGGCTACAGTTCAAATCATAGAACGCAGGCATGGT  
TCCATTTGTATCTCTTTCTTTTTTTGCTAATCTACCAAGTTATCTCTGATATGAA

AATTATATTGAGACTGACTAGTAACACAATCCTTTATTGTTACCAGGCAGAAAGTTT  
CAGCAGCAGAAGCAACCATTGAAGCCACTGCTCAAGGAATTGCCTCAATGCTTTGTG  
CTCATGGTCCTGAAGTTGAGTGGAGAATCTGCACTATATGGGAAGCTGCTTATGGAT  
TGATCCCTTTAAACTCCTCAGCGGTTGATCTTCCAGAAATCATTGTTGCTACCCAC  
TCCAACCTCCTATCTTGTTCATGGAATCTGTACATTCCACTCCTCAAAGTACTTGAGT  
ATCTTCCACGTGGGAGTCCTTCCGAGGCATGCTTGATGAAAATATTTGTTGCCACTG  
TGAAACAATCCTCAGCAGAACTTTCCACCAGAACTTCTAGGGATCATATCAGGA  
AAGCTAGATCGAGTTTAACCACGAGATCAGCGACCAAAAACCTTGCTATGGCTGAGC  
TTCGTGCTATGGTCCATGCTCTCTTTTTAGAAATCATGCGCTGGTGTGGAGATAGCTT  
CACGCTTGCTTTTTGTTGTGTTGACTGTATGTGTTAGCCATGAAGCACAGTCTAGTG  
GGAGCAAGAGACGGAGAAGCGAAGATGCTAGTAGCACTGCAGAGGAGAATCAACATG  
TGTCTGACAATCAAACCTAGTAACCGTAAAAGTAGGAGTGTCAAGGGACAAGGACCTG  
CGGCAGCATTGATTACATACGTTCTCGCTGCTGTGTGTGCTCTTGCCTGTGAGGTTT  
AGCTGTATCCTATGATCTCTGGAGGGGAAACTTTTCCAACCTGCAGTGGCTGCAA  
CCATTACAAAGTCTGTAAAGATAAATGGTTCATCTAACGAGTACGGATCTGGGATTG  
ACTCTGCAATTAAGCATAACGCCGTATCTTGGCGATTCTTGAGGCACTCTTTTCAT  
TGAAACCATCTTCTGTGGGGACTCCATGGAGTTACAGCTCAAGCGAGATAGTTGCTG  
CGCCATGGTTGCAGCTCATATCTCAGAACTGTTTCAAGCGTTCAAAGGCCTTGACGC  
ATGCGTTGTCTGGGTTGATGAGGTGCAAATGGGACAAGGAGATTCATAAAAGAGCAT  
CGTCGTTGTATAACCTCATCGATGTTTACAGCAAAGTTGTAGCATCCATTGTAGACA  
AAGCTGAACCCTTAGAAGCATAACCTTAAGAATGCACCTCTCAGAAGGATGCGCTGA  
CTTGTCTTAACTGGAAACAACAGAACACATCAGGGTTTGGTACAGCGGCGGTGATGT  
CCACGTCACGTAATGAAATGACTCCAAGAGGAGGTAACAACCATAAGTGTGCTAGAC  
ATTCAGATGAAGGGTCAGGGAGTAGATCCTCCTCCTCAGAGAAGGGTATCAAAGATC  
TGCTGTTGGATGCTTCTGATCTAGCGAATTTCTTACAGCTGATAGGCTCGCAGGGT  
TTTACCGTGGTACACAAGTGCTTTTTGAGGTCTATACTTGCGGAGAAACCTGAGCTTT  
CCTTCTCCGTTGTTTCGCTGTTGTGGCATAAACTGATCGCTTCTCCTGAAATCCAGC  
CCACCGTTGAGAGCACCTCTGCTCAGCAAGGATGGAGACAGGTAAAAGCCTTTCCCA  
TCTTGCTTACAGTTTACTTGATGTTTTTGTGGCTCTTCACGAGTCCAAGTCCTCAAG  
AAGCTAGACTTCATTCAGTATCATGTTTTCTCATGTATCAGGTAGTTGATGCACTAT  
GCAATGTCGTATCTGCAACGCCAGCAAAGCAGCTGCAGCCGTTGTTCTTCAGGTAC  
CTTATCATGAATAACACACCTGTGTAAAGATTAAGACCAATCAACTGCAGAAAACAT  
ATGGCTTATATAACCATGGTCTTGTTCCTCAAAGATTTACTTCTATGATCTAGTATGTT  
GAATGGTTGGTCTTGTGTTTAATATATACAGGCCGAGAGAGAGTTGCAGCCATGGAT

AGCGAAAGATGATGAAGAAGGTCAGAAAATGTGGAAAATAAACCAAAGGATAGTGAA  
AGTGATGGTGGAACTCATGAGAAACCATGACAGGCCTGAGTCACTGGTGATTCTGGC  
AAGTGCATCTGATCTTCTTCTGAGAGCAACTGATGGAATGCTTGTTGACGGAGAAGC  
TTGTACATTACCTCAACTTGAGGTAATAACACTATACACCTTCATAGATTTAAGGGT  
TTGCCCTTCCTCTGCCTAGACTTAAAAGCGTTTTTTTTCTTTGCTCTGACTTGCAGCT  
ACTTGAAGCCACGGCAAGAGCAATAACAACCAGTGTTAGCTTGGGGACCATCTGGACT  
AGCAGTGGTTGATGGTTTTATCCAATCTATTGAAGGTAAAGCATGAGCCTATCTGTGT  
TGATCAGTAAACAACTTTGGTGTGACGTTATGTGTGTGTTTTGTGATTCATTCTA  
GTGTCGTCTACCGGCAACAATAAGGTGCCTGTCACACCCAAGCGCACACGTACGTGC  
CTTGAGCACATCAGTACTACGAGACATCATGAACCAAAGCGGCATAACCACCGCCAC  
CACCAAGGCAACACCAAAACCACCGCCAACAATAACAACAGAGAAGAACGGAACGGA  
CAGTCCGTCCTACCGGTTCTTCAACGCGGCTGCAATAGACTGGAAAGCAGACATAACA  
AAAGTGTTTGAACTGGGAAGCTCACAGCTTGCTGTCCACGACAATGCCAACACAGTT  
TCTTGACACGGCGGCTCGGGAATTAGGATGTACCATATCAATGTCTGCCAATAACA  
CGAGCAGTACCTGCCTTCTTTAAAAATGATGTGGAAGTAGACCGACCAATACGACGT  
TGTACTTTTGTATTTAGTATATGTGTATGATGATGATGAGTGAC

**B.1.1.11. Protein sequence for DtGI (1179 letters)**

MTSPTSSEKWT DGLQFSSLLWSPPRDPQHKDQVVAYVEYFGQFTSEQFPDDIAELV  
RNQYPSTEKRLDDV LAMFVLHHP EHGHA VILPIISCLIDGSLVYSKEAHPFASFIS  
LVSPNSENDYSEQWALACGEILRILTHYNRPIYKREQQNGDTE SKASTSGSGSPTLS  
ETKVASPGQNERKPLRPLSPWISDILLAAPLGIRSDYFRWC SGVMGKYAAGELKPPT  
IASRGS GKHPQYMPSTPRWAVANGAGVILSVCDDEVARYETATLTAVAVPALLLPPP  
TTS LDEHLVAGLPAL EPYARLFHRY YAIATPSATQRLLLGLLEAPPSWAPDALDAAV  
QLVELLRATEDYASGVRLPRNWMHLHFLRAIGIAMSMRAGVAADAAAALLFRILSQP  
ALLFPPLSQAEGVEIQNSPLGGYSSNHRTQAEVSAAEATIEATAQGIASMLCAHGPE  
VEWRICTIWEAAYGLIPLNSSAVDLPEIIVATPLQPPILSWNLYIPLLKVLEYLPRG  
SPSEACLMKIFVATVETILSRTFPPETSRDHIRKARSSLTTRSATKNLAMAELRAMV  
HALFLESCAGVEIASRLLFVVLTVCVSHEAQSSGSKRRRSEDASSTAEENQHVS DNQ  
TSNRKRSVKGQGPAAAFDSYVLAAVCALACEVQLYPMISGGGNFNSAVAATITKS  
VKINGSSNEYGSGIDS AIKHTRRILAILEALFSLKPSSVGT PWSYSSSEIVAAAMVA  
AHISELFRRSKALTHALSGLMRCKWDKEIHKRASSLYNLIDVH SKVVASIVDKAEPL  
EAYLKNAPLQKDALTCLNWKQQNTSGFGTAAVMSTSRNEMTPRGNNHKCARHSDEG  
SGSRSSSSEKGIKDLLLDASDLANFLTADRLAGFYRGTQVLLRSILA EKPELSFSV  
SLLWHKLIASPEIQPTVESTSAQQGWRQVVDALCNVVSATPAKAAA AVVLQAERELQ

PWIAKDDEEGQKMWKINQRIVKVMVELMRNHDRPESLVIXASASDLLLRATDGMLVD  
GEACTLPQLELLEATARAIQPVLAWGPSGLAVVDGLSNLLKCRLPATIRCLSHPSAH  
VRALSTSVLRDIMNQSGITTATTKATPKPPPTITTEKNGTDSYRFFNAAAIDWKA  
DIQKCLNWEAHSLLSTTMPTQFLDTAARELGCTISMSSQ

**B.1.1.12. CDS for DtSOC1**

ATGGTGAGGGGAAAACTCAGATGAAGCGAATAGAGAATGCAACAAGCAGACAAGTG  
ACTTTCTCTAAGCGAAGAAATGGATTGTTGAAGAAAGCCTTTGAGCTCTCAGTGCTT  
TGTGATGCTGAAGTTTCTCTGATCATCTTCTCTGCTAAGGGAAAGCTTTATGAATTC  
GCCAGCTCCAATATGCAAGATACCATAGATCGTTATCTGAGGCATACCAAGGATCGA  
GTCAGCACCAAACCGGTTTCTGAAGAAAATTTGCAGCATTTGAAACATGAAGCAGCA  
AACATGATGAAGAAAATGAACAACCTTGAAGCTTCCAACGGAACTCTTGGGAGAA  
GGCATAGGATCATGTTTCGATAGAGGAGCTGCAGCAAATTTGAGCTACAACCTTGAGAAA  
AGTGTCAAATGTATTTCGAGCAAAAAGACTCAAGTGTTTAAGGATCAAATTTGAGCAG  
CTCAAGCAAAAGGAGAAAGCTCTAGCTGCAGAAAACAAGACGCTCGCTGAAAAGTGT  
GGATCTCATATGATCGAAGTTTGGTCAAATAAGAACCAAGAAAGTGGAAGAGGGGAC  
GAGGAGAGTAGCCCAAGTTCTGAAGTAGAGACACAGTTGTTTATTGGGTTACCTTGT  
TCTTCAAGAAAGTGA

**B.1.1.13. gDNA of DtSOC1 (black – non coding sequence, red – coding sequence)**

TTTAGAACAAAAAGAAGGAAGTGTGAGTTTTGTTTCTTGTTCCTTTTATTACACACA  
AATAGATGAAACGAGGAAAGTTATTACTTCTTTAGTTACTTGCATAAAAAGATCCTT  
CCTTTCTCAGTGAATCTAGTTTGATCATCTTCTCTCTCTTTTTTTTTCTCGTAGCTCC  
TCCAGTAATGCTTATATAGTCTCCTATATCTTCTTATTTATACACAATCCTGCTTCT  
TAGTTCACTCTTCTATTAATCTCCTTACAAGCTTCGTGGTTAGGTTTTCAATATTTA  
TTAATCCTTTTTGGCCACCGGTAAGATCTTTTCTTCATGACCTTTCTTGCTTTTTGCT  
TTCTAATCTTGGATTAGTCTGATAGATCTGCTTTCGTTTTTTATTTATTTGATGTTTT  
GTATCTGAGTTTTGTTATGCTTTTTTTTTTTGTTTCGGCTAAATCTCATGAAAGTTTGC  
ATATTTTCAGTTAATTCCAAGGACAAATTAACATTAAGTTGCATCTTTCATTTCGTT  
TTTAAGCAATCTCTATGTCACAACTAATAGATGTGTACTTTATACTGTCTTGCTTT  
TTGAGTACTTAATCATCTTCATCTCTCGTTGACTTAATCAACCTTCACACACACACA  
AAAAAGAAAAAATCAAATCCCAGGGGAAAATGGGTTCGTTTTATGTGTTGAGGTGGTT  
TGCACCACTAATCTCATGTTCTCTTCGATAAACCAAAAAAAAAAACAAAAATTGAAT  
AAATATGTTTTTTTTTTTACTGTGAAAGTGAATTAACAAGGAGAAAGAGATGGTGA  
GGGGAAAACTCAGATGAAGCGAATAGAGAATGCAACAAGCAGACAAGTGACTTTCT

CTAAGCGAAGAAATGGATTGTTGAAGAAAGCCTTTGAGCTCTCAGTGCTTTGTGATG  
CTGAAGTTTCTCTGATCATCTTCTCTGCTAAGGGAAAGCTTTATGAATTCGCCAGCT  
CCAAGTACGTTCTCTTTTATATATATCATCTACAGAGAAAGGGAGAGAGAGAAAAAA  
AGTTTTGTTTATATCCCTTGGTTTTAAATATGGATCTGTTGGCTGTTTATGGAATCA  
AATCTCTTTGACGTTTGCCTACTATTCGAGAGAGTGCTAAGTGAATTTCTTGGTTTC  
TACCACTGTAGGGATATTATTAAAGTAAACTAATTAAATCATCATGCTTAATCCTC  
TGATATGTCTCTCTCACATATTTTGTCTGTAATTATGATAAACTTAAATATCAA  
AGAACTATCTGTGAAATACTTTGCACTTGTATATATAAGATCATGATCACACACCC  
GTACACCTCCTTGAGTTTCTATTAAACACTTTCTTGAGTTCATTAGAAATAATAAA  
CATGTTGCAATTAGTAGTCTTTTGGTAAGAGATGATACTAAATCTGGTATAATTAAT  
CTGATTTGTTACTGCTTTAAGAAAAATCATATATTGCCATTAGTGCTGACTAGGAAA  
TTGTTTCACCAATTCTAATTAAAAATCATGCATGTTTCAATTCATGCTCATTCTTAG  
TATTATGTAGTTAAGTTGCATGTTTTGTTGCTATTGTTTCGGATTTTGCATCCAAAT  
CTCATGCCAGAATCTAAATGTGTAATAAGAAAGAAGTAATCTTATACTGTTAGGTAG  
AAGGGAAACCAAATGTACTTTTGATGAATTATTACATCTGTAAATAGAAAAAGCCAG  
CTTTTGATCAGATGTTTCAGACGGTTTGTGTACTTCTTTCTCTCAACCTAACTTAA  
TTTAAAGTAAAGaGACTCACATGTTTTTTGGTTTTCCAAAATGTTGGGGGAAATATC  
ACATATAGTACACGACTTTCTAAGTTGTGTCCTTAAGACTTGTGATATAACTTCACC  
TAAGAACTTTTGATGTTCTTATTTTATCTGTTTTTGAATCGAGAAATATTGATTTG  
GTTTCTTGTGTTTGTGCTAACTTAATCTGTAAATCATTTTCAATGTTTTAACAGTATG  
CAAGATACCATAGATCGTTATCTGAGGCATACCAAGGATCGAGTCAGCACCAAACCG  
GTTTCTGAAGAAAATTTGCAGGTTAAATATGATATTTTGTGCCATTATTAATCCAAC  
TCTAATATGAAATAAATTGGCTTAAGAAAAGTATCATATACAGTCTACATAATCTCA  
AGCCACATATTAAGGATAATTACATACTAACATTTTGTGTTATAACCAATAGCAT  
TTGAAACATGAAGCAGCAAACATGATGAAGAAAATTGAACAACCTTGAAGCTTCCAAA  
CGTTTGGTGATTCAATACAAACACATTATTGACTTGGATTATTGTATACATATCTTT  
AATAAGAGTTTGATCATATTGAGTATTTTCTTACTTTAATTAGGAAACTCTTGGGAG  
AAGGCATAGGATCATGTTTCGATAGAGGAGCTGCAGCAAATTGAGCTACAACCTTGAGA  
AAAGTGTCAAATGTATTCGAGCAAAAAAGGTGTATATATCATGCCACATCATTGTT  
CTTATAAAAGCCACTGAGATAAGTTCTAACTTAGTTACTCGTCTCTATTAGACTCAA  
GTGTTTAAGGATCAAATTGAGCAGCTCAAGCAAAGGTAATGAGACTCAAGTAGTGG  
AACACTCCAACCTTGATAATCTTCTTGTGCTTAGTTCTTACAATTGGTATAAAAAAAT  
CTTTGCAGGAGAAAGCTCTAGCTGCAGAAAAAAGAGCGCTCGCTGAAAAGGTACGTA  
CAATATATTCTATGGTTCTCTAGTTAGGGTTTCACCTGCACATCGTTTTCTTGTA

AACGAATATCTTAAATGGTTCTCAAGTTTCTGCACTTGCACATCGTTTTCTTTTATT  
CTTTTGTGATAGTTAACAGTTTGGTTCTTTTTTTTTGGGTTATTTCGATCGTCATGCT  
CATGACCTTGATCTATGTGTTTGTCTATCTTCTTTGCATCTATTAGTCTTGTGTAT  
ATATTAACAATAATTTTATTACAGTGTGGATCTCATATGATCGAAGTTTGGTCAAAT  
AAGAACCAAGAAAGTGAAGAGGGGACGAGGAGAGTAGCCCAAGTTCTGAAGTAGAG  
ACACAGTTGTTTATTGGTTACCTTGTCTTCAAGAAAGTGAACAAACGACTCTTCC  
AAAAACAAAAACAAAAACATAAAGCAAAATTTAAGTAAAAATGGGCTTAGCAGTGA  
ATGTACAAGTACAAGTAACAAAACCTCTAAACTGCTTCTGTTGTTGTCTTCTTTCTCT  
CTTCTTTAAA

**B.1.1.14. Protein sequence for DtSOC1**

MVRGKTQMKRIENATSRQVTF'SKRRNGLLKKAFELSVLCDAEVSLIIFSAKGKLYEF  
ASSNMQDITDRYLDRHTKDRVSTKPVSEENLQHLKHEAANMMKKIEQLEASKRLLGE  
GIGSCSIEELQQIELQLEKSVKCIKAKKTQVFKDQIEQLKQKEKALAAENKTLEAKC  
GSHMIEVWSNKNQESGRGDEESSPSSEVETQLFIGLPCSSRK

**B.1.1.15. CDS for DtTSF (plus 5'&3'UTR) (black – non coding sequence,  
red – coding sequence)**

GATTGGTTGAAAGTTTAGATCAAATATCATGTCTTTAAGTCAGAGAGATCCTCTTGT  
GCTTGCAAGAGTTGTGGGTGATGTTGTTGATCCTTTCATGAGATTGATCTCTCTTAG  
AGTAACTTATGGCCAAAGACAGTTATAAATGGAATGGATCTAAGGACTTCTGAAGT  
TCTAAATAAACCAACGGTCGAGATTGGAGGAGATGACCTCAGAAATTTCTACTTTT  
GGTTATGGTGGATCCAGATGTGCCAAGTCCAAGCAACCCTCACCTCCGAGAATATCT  
CCATTGGTTGGTGA CTGATATACCTGCAACA ACTGGAACAGCTTTTGGAAAAGAGTT  
GGTGTGCTATAAGAATCCATGTCCCACTTCAGGAATTCACCGAGTTGCATTGATGTT  
GTTCCGACA ACTCGGAAGACAAACCGTTTATGCACCCCAGTGGCGCCACGCTTCAA  
CACTCGTGAGTTGCTGAGAACTACAATCTTGGTCTACCCGTTGCTGCCGTCTTCTT  
CAACTGTCAGAGAGAGAATGGCTGCGGAGGAAGAACAATAGAAAGAGTATCCATCC  
CTCGTTCCTCTTAAGTAATATCCATTTAAAGGTTCTTGTGTTGTAATAAGAAATATAAT  
GATTATGATATGACTAACATAAGGTGACGCATGCATGCGGACATCTATATAAAGTA  
TATGTAATAACTAATAATTGAGAGGGCCGAGGAATGAAAAATGAGTGTGTTCTACTT  
GTTAAAAA AAAAAAAAAAAAAAAAAA

**B.1.1.16. Protein of DtTSF**

MSLSQRDPLVLRVVGDVVDPFMRLISLRVTYGQRQVINGMDLRTSEVLNKPVEIG  
GDDLRFNYTLVMVDPDVSPSPNPHLREYLHWLVTDIPATTGTAFGKELVCYKNPCPT

SGIHRVALMLFRQLGRQTVYAPQWRPRFNTREFAENYNLGLPVAAVFFNCQRENGCG  
GRRT

## Appendix C.

### C.1. SNPs identified in whole genome sequencing

#### C.1.1. SNP 1) 85295 C1\_0 C2\_18 C1\_14 C2\_0 A/G

ttctatgtatctttaaagctcatcttgggtcaaccacaaacagcttttagaaagata  
taagtttagacgagtatatttagttcttgtaatcggccgacaatttatgcatatgat  
attaatacaaaattatccggatcctttacatattttataacattatattttacat  
caatgaccggatgatgcaatccacagtaggtacatcaatgatcagaagtctcaacat  
cagagcctccttgttcatttttctaaagttggataattatgacatgctgcagat  
ttgtaaacttttctctataaatatgcataatttctttatgtgggacttgaaaccaa  
gcctccttatgtagAATTATTAATGCACCTTTTATCAAACCACTA/GAACCAAGAAG  
GCTTCCACTAGCCTCCTTGTtcagaccattattgtttcttgttagtgttaaatcaat  
gttactatttacaatgaggaaaccgaaaacaatgataactagaaccataaacat  
aagttcattgggaaaccataaacaatttaaaaaccagattaaacccctcaaaagcat  
agttcaataaaaccaagataaacaagtctgtataatcgagatTTTTTTctgcggaat  
caaaaataaaatggcatttacttatcttggtttcatcatattgtatactatcttaca  
caactaaaactgctatgtaacctagaaagtaaattaatggcgaagTTTTTctctct  
gacaacgccgcaaatttgacagggctgtgttccagggtt

#### C.1.2. SNP 2) 76246 C1\_8 C2\_0 C1\_0 C2\_9 A/G

ttgattcagtcataatcttgggtatcaaaacagaaaaagttaggcatgaacaaaat  
ctttttttttcttttgggtgaagatcttgataacaaagtcggtgttcacattttgtg  
gtatagtttttagtaacatgaatccatgtacatgtatcttgggtatagttttgtgatcaa  
agcagagtgttttcccatgttatctactataaacacagtcacacattactcattac  
ttatccaacatcacccgtccatgtatatcttgtctgcaaatcatccatccatgcatc  
accatctacacatcaccttaataatatctacaattgttgtttctattaaggggaaa  
ttgtccacattatgtggctgaccaacaataaggtgtatcacatgtaaaaagtgtttt  
taagtgtaaaaaaaagtcaaaaagtgctcgcttagtgtaataaactctagtttgata  
ggtaatgaatcttggtttctacttctcccctcttgggttttgcatcttgccttat  
ctccaaaagatctatgctgttaacttcttaaatagcttatcatcactagcttttc  
agatgcactaaattatccacgtgtaatgatctttatccacgtctaaggccattcgca  
gtggtcacatcttactaggctacctgacgataaaaactattatataaattagaaca  
agaaatagaaaacaagaaatgagagaaacgtgtccgtggaaaaatatgttttatttg  
atctgagagggttaattctacagatgtcagcgccacatgtttccatactttaaatatt

aatattttaattattttatcttaaaataactaatatataactatagagacgttggtaa  
aGAAATGCACTAATAATGATGCTTTAACTCCA/GATGTCACTTTTCTACATCATAAA  
TCTAATActataaagaatagtataaga

**C.1.3. SNP 3) 77982 C1\_0 C2\_8 C1\_12 C2\_0 C/T**

tgtagggttgttctcattggttccccattgtagttgcctctctttttgcattcttaat  
gatcagactcaatgtcaaaagaaatagtttcagtttctgtaattatggcagttcatt  
atgcttacgatcagctagagtcgatctttacaaggatggtgatcatggcatagtaac  
aatgcctcttaaagctagtaaataagaaaatattattgaagcagaaagtgaccata  
gccttgagataatctcgtcaacaataattcctaaaccattgaggtcaaaccacag  
cgctccatttcgccaccaaaaatttattttgtatcactcagttcaactcttcca  
tactgtcaagaacatctgggagcatcctcaggcttttgaaacctcgcatgctcaatt  
tagcaagatctggccaagaccttacagatgatggaacttcttctataaccagtatttt  
ccagatccagatagccaattcttgtagaaatctctggaaacgttttcagttgtgtgc  
agtgactgaggtcaagatgatataatagagactccatggtgatggttggttggagaac  
cttgagattcttgcacccaccatcttaagggtgtgtcagacttgagagattctggat  
ccaagaaggaaatctccaccaGACTTTCACAAGCATTGAGTACTAATATCTC/TAAGA  
TTCTTGGCATTGACACATTAGGGACgtctttcaagtttaaggagtaagacatatcc  
atcagcttgagatgtcgaagccgctgtacatatggaaagtggaaatagacaaatcaa  
aacacatatataatggtgagcaaattctccaaaagaaattgaaataagtttttttt  
taccaaaatataaaatggctaaaaagtttattcaaaacagtagaagatcattttac  
atttgttt

**C.1.4. SNP 4) 141907 C1\_0 C2\_8 C1\_8 C2\_0 A/G**

ctaatacgaagcataagcatagtttcatttcgtctggccagaaaaatgaagattttg  
ttcctaacaggaattgaacttaaaatctatggactgtatatacattatcccaaattt  
tatttttaccattagattaccaccggttaccggttgattacattatcaattattact  
ttcttatgacaagtgataaaatttatagataaataatacaagtagtattcactttct  
cttacaatggatataatattatgtaaatggtactaaaatgtaacatcaagatagaa  
aatagtggtatttagaggagaaagtgatctggttcaaaaaaagaggagaaagtgaag  
aatggcaaggcttacttttaattgctttgcttcaaaaggagatcctttgtagatct  
aatgtatgatatcgtaagtacgttataataactcagattaggtacaatagacacaaaa  
ctaatacgaagggacagaaagaccgatttaagtatacatatataatttttagtcaaa  
aaaaagagtatacgtatataatataattataacttataatttctaaagaagtgacta  
agcaaaatacttgtaacgtaacgtacgttggttaaactaacataacattaagcagtaa

ataaatatgattgtcaatgcttatttttagtCTGGTAAAATATTGTAGTAAGATCTTT  
CTCA/GAGCAAACATGAAATATACTAATGGTTGGAAggatggctagtcctctggaac  
acataaaaagaaatatgccttgttactatttctattttatgtcgaatattggagtt  
tatgtaaagggtggt

**C.1.5. SNP 5) 63435 C1\_0 C2\_9 C1\_13 C2\_0 C/T**

ttgtacttggctcgatcaactgaaagaattccaaattcaaaaaaattggaattccac  
agttacttgtaagggttttattgatttgtgtattgtgtttaaaaatgcactaaggcct  
tagaccttagcattcATATTGAATTTTCAGCTTTTCTGAGTTTATC/TTATATAAC  
TAACTTATATTATAAAGAAGTCcacataccacttaaaacgcccttggttttttttctt  
aatacaaatgtatgtgtgggtttttatcttcagatactgatgtaggtaaaatgggt  
gatgatcttctggtgaattctaattacctcgcaacagaaagaactacttttaatgct  
tcagagagtaaagtgggtgtgttttgggatataattgggtgtcccatggccattaga  
cattggaaagatattcctataatcatcaaaactgctctggaaatggcaggatacatg  
agtgagctagaaat

**C.1.6. SNP 6) 109293 Q1\_72 Q2\_0 , A/G rank\_0.81507,**

ttcttcactaaactaatgatatttgataaattaacttatactgaagaaaaagaaaa  
tttgtagggcgccacatccggttcttggtaatcgcatctcatatgtccacatgttttta  
actaccaaagtatcagttatcaaccgcaattctttgatttgatacatgcaagtcaa  
tctaagtttttttttgctcgctaacgtcaaactttgtgtttcacaatccagtatg  
aacacaaaaactagcacatactgtattgttttgaagtctaattaatatataagctg  
atattttaagcatatctacacatgaaattaaccgCCAAGTTACCTTAAAATATCCAT  
ATAGAGGA/GTTCAAGTACTGGTACTACTATATGTTAATAcatcgggtattgtctat  
aatacgaatatcagttatgtatgtgtatatatttacatcagaagacaaaatatat  
ccatcaaatattaaatagactgtaacaagaaatccaagtttattcaaatcacaac  
agtacaacacacacatatataaacatgatactatagagctgcttcccgttattaaga  
tgtaga

## Appendix D.

### D.1. Protoplast isolation results

Arabidopsis	Method	Number of cells	Viability	Transformed	Rockett	method	Number of cells	viability	Transformed
1/5/14	1	300,000	n/a	no	1/5/14	1	164,000	n/a	no
7/5/14	1	242,000	n/a	Gfp, YFP+crispr - failed	7/5/15	1	64,000	n/a	Gfp, YFP+crispr - failed
					13/5/14	1	258,000	n/a	Gfp+crispr - failed
14/7/14	1	228,000		Gfp+crispr	14/7/15	1	270,000	n/a	Gfp+crispr
19/5/14	1	210,000	n/a	Gfp+crispr	19/5/14	1	142,000	n/a	Gfp+crispr - failed
	1	204,000	n/a	Gfp+crispr		1	154,000	n/a	Gfp+crispr - failed
					21/5/14	1	266,000	n/a	Gfp, YFP+crispr
27/5/15	1	238,000	n/a	gfp	27/5/15	1	264,000	n/a	gfp
					28/7/15	1 (0.6M)	224,000	n/a	YFP
					11/8/14	1 (0.65M)	230,000	12.35	no
					11/8/14	1 (0.7M)	230,000	15.14	no
					13/8/14	1 (0.6M)	388,000	20	Gfp+crispr
					22/8/14	1 (0.4M)		30	
					26/8/15	1 (0.6M)	242,000	1 or 2 cells only	
					26/8/15	2	462,000	37	Gfp+crispr
					9/9/14	1	350,000	21.37	

					11/9/1 4	2	262,000	30	
					15/9/1 4	1 (0.4M)	228,000	11.5	
					15/9/1 4	1 (0.6M)	201,200	15	
					17/9/1 5	2	266,680	3.5 (new pectina se)	Gfp+crispr
					22/9/1 4	1	275,000	3.8 (17.3 sterile)	Gfp+crispr
					25/9/1 4	2	246,680	3.5 (new pectina se)	

## D.2. Primers used in targeted mutagenesis experiments

### D.2.1. CRISPR *FT* target primers

CRISPR\_FT\_1\_F – GATTGGTACTTATGGCCAAAGAG

CRISPR\_FT\_1\_R – AAACCTCTTTGGCCATAAGTAACC

CRISPR\_FT\_2\_F – GATTGCTCGGAGGTGAGGGTTGCT

CRISPR\_FT\_2\_R(2nd) – AAACAGCAACCCTCACCTCCGAGC

CRISPR-FT-96(n102)96-98-For – GATTGTCACTCTAAAGGTTACTTA

CRISPR-FT-96(n102)96-98-Rev – CTAAGTAACCTTTAGAGTGACAAA

CRISPR\_FT-136(133-155)\_F – GATTGGCTTGTTTTGAACCTGAGA

CRISPR\_FT-136(133-155)\_R – AAACCTCTCAGGTTCAAACAAGCC

### D.2.2. Primers for the isolation of *AtFT*

AtFT\_5'UTR\_1\_F – CACAGAGAAACCACCTGTTTGTTC

AtFT\_3'UTR\_1\_R – AGGCATCATCACCGTTCGTT

AtFT\_full\_1\_F – AGATGTCTATAAATATAAGAGACCCTCTT

AtFT\_full\_1\_R – GAAGCCATCTAAAGTCTTCTTCCTCC

AtFT\_seq\_1\_R – GTGTAAGAAGCACTGAAGAACTA

AtFT\_seq\_2\_R – GTTCCAGTTGTAGCAGGGATATC

AtFT\_seq\_1\_F – GCATAAGTGATTATTATTTAGGATAA

**D.2.3. Primer pairs used in PCR for cloning of shorter *AtFT* fragments**

**CRISPR FT-1** – AtFT\_full\_1\_F – AGATGTCTATAAATATAAGAGACCCTCTT

X AtFT\_seq\_1\_R – GTGTAAGAAGCACTGAAGAACTA

**CRISPR FT-2** – AtFT\_seq\_1\_F – GCATAAGTGATTATTATTTAGGATAA X

AtFT\_seq\_2\_R – GTTCCAGTTGTAGCAGGGATATC