Analysis of the molecular mechanism regulating anther endothecium secondary thickening in *Arabidopsis thaliana*

Stefan Fairburn[1,2]

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2019

Supervisors: Professor Zoe Wilson[1], Professor John King[2]

[1]Plant and Crop Sciences, School of Biosciences, The University of Nottingham
[2]School of Mathematics, The University of Nottingham

Abstract

Food security is important to feed a growing global population in a changing climate. Heterosis (the phenomenon of increased yield in hybrid crops) is a significant way to maximise crop yields without increasing input. One difficulty with breeding hybrid plants is preventing self-fertilisation, and there is therefore interest in male sterile plants in commercial crop systems. This can be achieved by controlling development or release of pollen.

MYB26 has been shown to be a key driver of anther endothecium secondary thickening, which is critical for anther dehiscence, working via NAC transcription factors *NST1/NST2*. Here mathematical modelling is used to suggest that *NST2* upregulates *MYB26* at the post-translational level. Other anther endothecium genes are investigated with the F-box protein SAF1 being shown to be downregulated by MYB26 and seeming to negatively regulate NST2 accumulation. A potential network using these observations is hypothesised and further equations to test out this hypothesis in the model are suggested.

Possible *SAF1* orthologues and redundant genes are identified and investigated. Knock out lines of these genes were investigated to determine their role in the network.

TGA9, and its orthologue *TGA10* have been shown to encode proteins which interact affect MYB26 and double knockout lines to be anther indehiscent. Their role in relation to SAF1 is investigated.

Acknowledgements

I would like to thank both of my supervisors, Professor Zoe Wilson and Professor John King for their invaluable help throughout my Ph.D. project and whilst writing my thesis. Professor Wilson has been very important in my personal development along with helping me develop critical scientific thinking. Additional thanks should be given for giving me the confidence and opportunity to spend 3 months working in China. Professor King has been key in my development of mathematics and the combination of maths with biological, real world systems.

I would like to thank all of the members of Professor Wilson's laboratory, with particular thanks going to Dr. Alison Tidy who was crucial for my development of scientific and molecular techniques. A huge thank you must be extended to Dr. Leah Band who was invaluable with regards to modelling and providing coding to simulate networks. The whole group has been incredibly friendly and I have thoroughly enjoyed my time working with and alongside everyone in the lab, and in the wider Plant and Crop Sciences department. Again, extra thanks should be given to the support staff within the department, particularly Miriam Colombi, who was crucial to my 3 month internship in China.

Thanks should be given to my funding body, the BBSRC who have funded me and my Ph.D. project, allowing me the opportunity to gain a Ph.D.. Thanks to The Centre for Plant Integrative Biology, particularly their Net4FS programme, which funded my internship in China.

Finally, I would like to extend my deepest thanks to my family, friends and my girlfriend who have had to put up with me for the last 4 years and have been a constant source of support throughout my Ph.D..

Contents

List of Figures

List of Tables

List of Abbreviations

bp: base pair

CaMV: Cauliflower Mosaic Virus

cDNA: complementary Deoxyribonucleic Acid

ChIP: Chromatin Immunoprecipitation

Dex: Dexamethasone

DNase: deoxyribonuclease

dNTP: Deoxynucleotide triphosphate

dATP: Deoxyadenosine triphosphate

EMSA: Electrophoretic Mobility Shift Assay

FRET: Förster resonance energy transfer

g: gram

GFP: Green Fluorescent Protein

GUS: β-glucuronidase

HCl: Hydrochloric acid

Kb: kilobase pair

LB: Luria Broth

min: minute

ml: millilitre

mM: millimolar

mRNA: messenger Ribonucleic Acid

MS: Murashige and Skoog Basal Medium

NASC: Nottingham Arabidopsis Stock Centre

PCD: Programmed Cell Death

PCR: Polymerase Chain Reaction

PMI: Pollen Mitosis I

qRT-PCR: Quantitative Reverse Transcriptase-Polymerase Chain Reaction

RT-PCR: Reverse Transcriptase-Polymerase Chain Reaction

ST: Secondary Thickening

TAIR: The Arabidopsis Information Resource

T-DNA: Transferred-Deoxyribonucleic Acid

UTR: Untranslated Regions

Wt: Wild type

Y2H: Yeast-2-Hybrid

YFP: Yellow Fluorescent Protein

Chapter 1: Introduction

1.1 Food Security

Global food security is one of the major concerns facing scientists at this moment in time. Food security is defined by the Food and Agriculture Organization of the United Nations (FAO 2014) as being when "people, at all times, have physical, social, and economic access to sufficient, safe, and nutritious food that meets their dietary needs and food preferences for an active and healthy life" (Pinstrup-Andersen 2009). Currently there are 2 billion people who are not 'food secure' (FAO 2014) and this is only going to continue get worse as populations continue to rise and diet preferences change (Godfray *et al*. 2010). This problem is exacerbated by a number of other factors, including climate change which has had a number of potential negative effects on agriculture ranging from more unpredictable weather patterns (Gregory *et al*. 2005) to an increase in plant disease (Coakley *et al.* 1999), or increased losses due to pests (Deutsch et al. 2018), all which are likely to only continue to get worse as global temperatures continue to rise.

1.2 Heterosis and Male Sterility

Due to the increased demand for food, coupled with the limited land and water available for food production (growing populations also require increased habitable land and some land is not cultivable), it is important to maximise yields. One way to achieve this is through heterosis – the fact that hybrid plants produce higher yields than their parents (Duvick 1999; Lippman and Zamir 2007). Hybrid plants are plants grown from two different parent varieties. For hybrid plant growth to be an option, there is need to prevent self-fertilisation within the same variety. This is usually achieved by emasculation (removing anthers) of plants, however achieving this is challenging with either the need for application of chemical gametocides, which are difficult to apply and frequently unreliable, or the use of male sterile lines which are limited by

germplasm and the need for fertility restoration systems (Wilson and Zhang 2009). Previously the focus has been on pollen viability and production with regards to producing sterile lines (Feng and Dickinson 2007; Ma 2005), however this may lead to issues when trying to maintain lines of these crops due to difficulties is rescuing fertility (Wilson *et al*. 2011). One option to overcome this is to produce plants which are male sterile because they fail to release pollen, yet still produce viable pollen. This can achieved by affecting anther dehiscence (Wilson *et al.* 2011).

1.3 Anther and Pollen Development

1.3.1 Anther Development Overview

During plant floral development, pollen is formed within the stamen of flowers. These are organs which consist of a long stalk known as the filament, which has a wider region at the top where pollen is formed – the anther. In Arabidopsis floral development is separated into 12 stages, with the anther emerging at stage 5 (Smyth *et al.* 1990). At this point the stamen primordia emerge from the flowering bud in the third whorl (Goldberg *et al*. 1993) and then develop into the stamen – the anther and the filament – through cell differentiation. Anther development is separated into 14 stages, and subsequent references to stages of development refer to anther development stages (Sanders et al. 1999) and development of the stamen in the developing flower is illustrated in Figure 1.1. During phase 1, anther morphology is established, cell and tissue differentiation occur, and meiosis of the microspore mother cell occurs. Most specialised tissues and cells are present, along with the microspore tetrads within pollen sacs. Phase 2 involves the differentiation of pollen grain, enlargement and dehiscence of the anther, and eventually pollen release occurs.

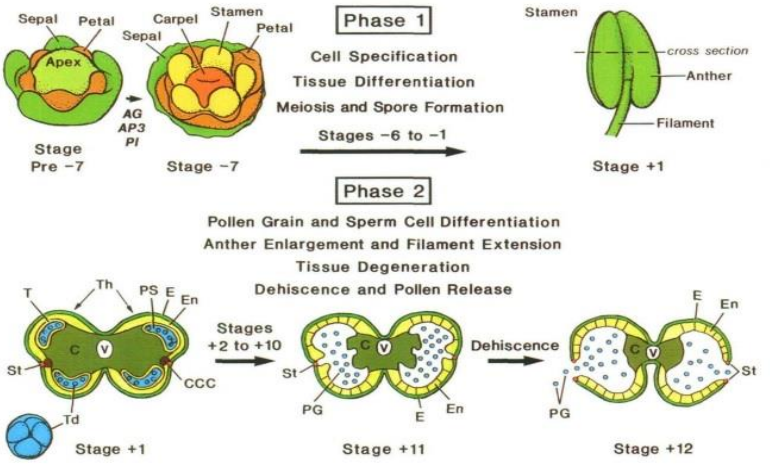

Figure 1.1 – Schematic representations of anther developmental stages and cross-sections from Goldberg et al. (1993). C, connective; CCC, circular cell cluster; E, epidermis; En, endothecium; PG, pollen grain; PS, pollen sac; St, stomium; T, tapetum; Td, tetrads; Th, theca; V, vascular bundle.

The first stages (1-5) establish the shape along with the cellular differentiation of the four locule cell layers in the developing anther. This involves divisions of the hypodermal cells in the anther primordium into four clusters of archesporial cells in the anther lobes, which then divide again to produce two cell layers – the primary parietal (PP) and the primary sporogenous (PS) cell layers. The PS will go on to form the diploid pollen mother cells (PMC) which then develop into pollen, whilst the PP forms the maternal cell layers surrounding the PMCs (Scott *et al*. 2004; Wilson *et al*. 2011). These are, from the layer closest to the PCM outwards, the tapetum, the middle layer and the endothecium (Li *et al*. 2006). This is stage 5 and it is at this point where the anther has developed its 4 lobed morphology. Once the PCM and the cell layers have developed the PMC undergoes meiotic divisions at stage 6 to form four haploid microspore cells (Ma 2005; Stern *et al.* 2011). The meiotic cells differentiate from each other and the tapetum, giving rise to a cavity inside the tapetum called the locule, where the microspore cells are located. Stage 7 is defined as the point where meiosis is complete. Whilst this is occurring the tapetum secretes nutrients which are

taken up by the microspore cells, along with lipids, enzymes and polysaccharides which are used to build the outer wall around the developing pollen cells (Mariani *et al.* 1990; Piffanelli *et al.* 1998; Shi *et al*. 2011). This development from microspores to pollen takes place during stages 9-12. Once the pollen has matured, at stage 12, mitotic divisions occur which leads to tricellular pollen grains which are haploid. Alongside this, at stages 10 and 11, the tapetum is broken down whilst the endothecial layer undergoes selective secondary thickening, in anticipation for anther dehiscence, which occurs at stage 13 when the stomium degrades. This anther dehiscence can be coupled with an elongation of stamen filaments in some plant species, for example wheat plants (Kirby 2002). Stage 14 involves the senescence of anther cells after pollen release in the previous stage, with anthers eventually detaching from the filament. (Ma 2005). Because of the importance of pollen and anther development in reproductive success across plant species, the general developmental pathways are conserved very strongly across species *(*Gómez *et al*. 2015).

1.3.2 Anther Dehiscence

Anther dehiscence is the process at the end of pollen development which involves the opening of the anther to allow mature pollen to be released. It is of particular interest because by controlling dehiscence of plants it is possible to cause male sterility without affecting pollen development – this means it is easier to recover fertility because viable pollen is produced. For example, plants with reversible male sterility have been generated by mutating genes involved the jasmonic acid biosynthesis pathway causing male sterility, with exogenous application of jasmonic acid able to recover fertility (Park *et al.* 2002). A number of male sterile mutants have been identified with defects in anther dehiscence and previous studies have characterised key genes involved in regulating this process. For anther dehiscence a series of coordinated events are required.

Two specific specialised tissues are especially important in opening as part of anther dehiscence – the stomium and the septum of developing anther (Wilson *et al.* 2011). The stomium is made up of specialised, modified cells within the epidermis and determines the position of anther dehiscence, whilst the septum is the cluster of cells separating the 2 lobes within the anther (Figure 1.2).

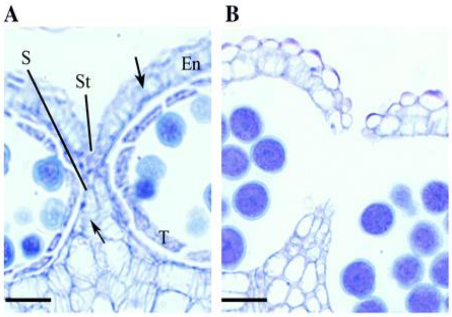

Figure 1.2 – Developing anthers photographed to highlight the septum and stomium in wild type *Arabidopsis.* (A) displays the stomium region following release of the microspore whilst (B) shows this region at the moment of septum and stomium lysis – anther dehiscence. Secondary thickening of the endothecium can be observed in (A), highlighted by the arrows. En, endothecium; St, stomium, S, septum.

Degeneration of specific anther tissues is required – particularly the stomium and septum – which occurs at stage 13 (Sanders et al. 1999). It has been established that Programmed Cell Death (PCD) (Keijzer 1987; Wu and Cheung 2000) of tapetal cells is required for normal development of pollen cells (Kawanabe *et al.* 2006; Parish and Li 2010), and a similar PCD mechanism is thought to be important in the breakdown of stomium and septum tissues in anther dehiscence. Use of mutant analysis has identified a number of genes important in the role of degeneration of the appropriate tissues in developing anthers (Sanders *et al.* 1999; Scott *et al*. 2004), for example *NON-DEHISCENCE1.* Mutant plants for *non-dehiscence1* have a failure of stomium region breakage due to abnormalities in PCD, and so were male sterile despite developing normal, viable pollen because the pollen was trapped within the anther (Sanders *et al.* 1999). Currently, it is believed that PCD commences in developing tapetal tissue and spreads out into the middle cell layer and the stomium (Varnier *et al.* 2005). The actual process of breaking down cells has

been suggested to be mediated by cysteine proteases (Xu and Chye 1999), although additionally there is evidence that Reactive Oxygen Species (ROS) may also play a role (Dong *et al*. 2005; Dai *et al.* 2019).

Equally as important to normal anther dehiscence is the selective thickening of the anther endothecium. There are some differences in types of anther endothecium thickening, with 4 main types having been observed in a generally species-specific manner (Wilson *et al*. 2011), however role of this secondary thickening is concerned. The endothecium undergoes thickening from stages 6-10 of anther development, with selective secondary cell wall thickening then occurring at stage 11 (Sanders *et al*. 1999; Scott *et al.* 2004). This secondary cell wall thickening has been shown to consist mostly of lignin and cellulose through phloroglucinol and ethidium acridine orange staining (Dawson *et al.* 1999; Yang *et al.* 2007), with the composition of this secondary cell wall also being important in anther dehiscence (Thévenin *et al.* 2011). Selective secondary cell wall deposition leading to thickening of the anther endothecium (which does not occur in other tissues) leads to the sections of the endothecium being stronger than surrounding tissues.

This enzymatic breakdown of the stomium and septum, along with the selective thickening of the anther endothecial tissues is combined with the dehydration of the endothecium and epidermal cells. This has been suggested, in part, to occur due to stomatal evaporation of water (Keijzer 1987), however there appears to be active relocation of water within anthers and petals (Bonner and Dickinson 1990). Observation of starch levels and H+-sucrose transporter, AtSUC1 localisation in the connective tissues of developing anther suggest that an increased osmotic pressure could be used to draw water out of the anther (Stadler *et al.* 1999). There is also an accumulation of cation localisation observed in developing anthers which could also be used to increase osmotic pressure around the anther leading to dehydration (Matsui *et al.* 2000; Rehman and Yun 2006). Aquaporins *PIP1* and *PIP2* have also been identified as being

anther and style specific, and likely allow the passive movement of water across membranes out of the anther during dehydration (Bots *et al*. 2004; Bots *et al.* 2005).

The combination of the breakdown of stomium and the septum, the secondary thickening of the endothecium and the dehydration of the anther synchronously occurring leads to differential pressures on the anther. The dehydration of the anthers causes the locule to bend outwards. The strengthened endothecium does not break under the forces exerted on it by this dehydration, whereas the weakened stomium and septum can break open releasing pollen from the anther. It has been shown that these differential pressures are mathematically enough to lead to anther dehiscence in *Arabidopsis* by developing a model describing the biomechanics of all the various cell layers and how forces interacting on different tissues would be enough to lead to rupture of the relevant cell layers (Nelson *et al.* 2012).

1.4 Regulation of Anther Endothecium Thickening

1.4.1 Genetic Control

1.4.1.1 MYB26/MALE STERILE35 (MS35)

A number of genes which affect the jasmonic acid (JA) pathway have been identified as leading to male sterility when mutated. These are mainly in water transport during the dehydration of anthers as dehydration is a process which seems to be initiated by increased jasmonic acid accumulation (Ishiguro *et al*. 2001; Park *et al*. 2002; Sanders *et al.* 2000; Stintzi 2000) (further information about anther dehydration in Section 1.3.2, further information about JA in Section 1.4.2.1). However, a JA independent gene, *MYB26*, has been identified as having a major role in endothecium secondary thickening (Yang *et al.* 2007). The *myb26* mutant was isolated from a collection of mutants generated by x-ray mutagenesis of *Landsberg erecta* (Ler) seeds and homozygous plants are male sterile due to anther indehiscence – pollen is normal and viable and plants

are female fertile (Dawson *et al*. 1999). MYB26 is a R2R3-type transcription factor, which are thought to act as transcriptionally repressors or activators.

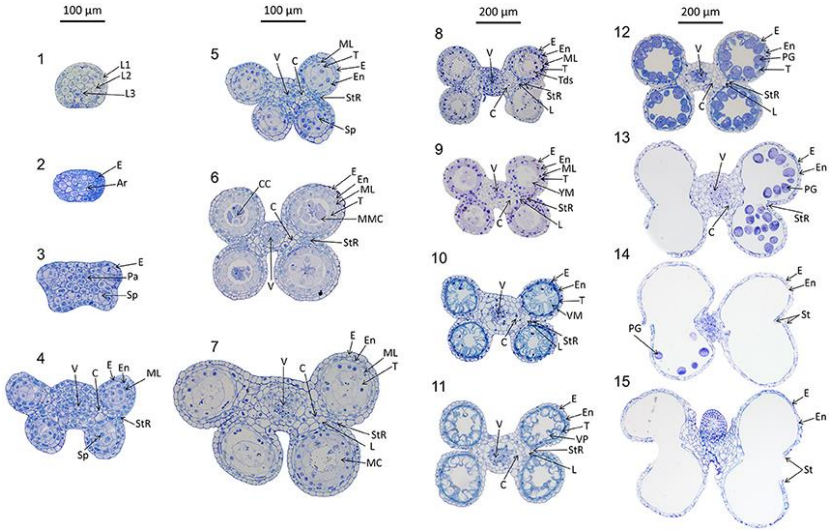

Figure 1.2 – Stages of anther development in *Triticum aestivum* as displayed by Browne et al. (2018). Although this is a different species, *Arabidopsis* anther development is the same and this is a excellent diagram of anther development. L1, 1st Cell layer; L2, 2nd Cell Layer; L3, 3rd Cell layer; Pa, Parietal Tissue; Sp, Sporogenous Tissue; E, Epidermis; En, Endothecium; ML, Middle Layer; T, Tapetum; L, Lacunae; StR, Stomium Region; MMC, Microspore Mother Cells; Tds, Tetrads; YM, Young Microspores; VM, Vacuolate Microspores; VP, Vacuolate Pollen; PG, Pollen Grains; MC, Meiotic Cells; V, Vascular Region; C, Connective Tissue; CC, Central Callose. Scale bar is 100 µm for stages 1–7 and 200 µm for stages 8–15.

MYB26 mutant and wild type plants are developmentally equivalent until stage 11 of anther development (Figure 1.2), when *myb26* plants have no lignification of the endothecium and subsequently pollen release fails. Lignification of non-anther tissues is unaffected in the *myb26* plants, which may suggest that *MYB26* has anther specific expression. One could argue that this point is countered by the fact that GUS staining shows strong expression in the styles and nectaries of ProMYB26:GUS plants, although there was no significant difference in phenotype between wild type and *myb26* plants (Yang *et al.* 2007)

and MYB26 protein has only been seen in the endothecium layer, however it may be that *MYB26* is expressed in non-endothecium tissues and the protein is transported into the endothecium, or alternatively expression of *MYB26* in non-endothecium tissue occurs but MYB26 protein is degraded before it can have any effect. The secondary thickening role of MYB26 is highlighted by ectopic secondary cell wall thickening throughout the plant when *MYB26* is overexpressed.

Expression of *MYB26* has been shown to be highest during pollen mitosis I before dropping away later in development. This corresponds to the time before secondary thickening begins, suggesting that MYB26 is important in earlier stage development of endothecium thickening. This may suggest that *MYB26* leads to secondary thickening of the anther by turning on a number of other genes which have to accumulate to synthesise secondary cell wall thickening. Additionally, as mentioned earlier, overexpression of *MYB26* leads to ectopic deposition of secondary cell walls in other tissues, strengthening the suggestion that it plays a role in the thickening of the anther endothecium (Yang *et al.* 2007).

Investigation of the expression of other genes in relation to *MYB26* in the *myb26* mutant and overexpression lines has been previously carried out (Yang *et al.* 2007). It is known that two NAC factors *NAC SECONDARY WALL–PROMOTING FACTOR1* (*NST1*) and *NST2* have an impact of endothecium thickening due to previous studies (Mitsuda *et al.* 2005). Originally it was suggested that these two NAC factors regulated the expression of *MYB26*, however data suggests that this relationship is more complex, with *NST1* and *NST2* expression reduced in the *myb26* mutant and overexpressed in the *MYB26* overexpression lines (Yang *et al.* 2007). Further work by Yang *et al.* (2017) showed that induction of *MYB26* expression directly upregulated *NST1* and *NST2* expression. However, overexpression of *NST1* or *NST2* did not rescue the indehiscence phenotype in *myb26* lines, and overexpressing *NST1* and *NST2* in this background led to

ectopic secondary thickening in the epidermis, but not the endothecium. This suggests that *MYB26* plays a role in the expression of *NST1/NST2*, but that the lack of ectopic expression (as visualised with GUS promoter systems) in the anther endothecium in *NST1* and *NST2* overexpression lines may be the result of a negative regulation of *NST1/NST2* specific to the anther endothecium.

The fact that *MYB26* is expressed earlier than secondary thickening occurs, coupled with the fact that *MYB26* expression correlates with *NST1* and *NST2* expression could suggest that MYB26 acts as a master switch for endothecium thickening. Downstream of *MYB26* are *NST1* and *NST2*, with *NST2* being particularly important in anther development since it is more sensitive to *MYB26* expression (Yang *et al*. 2017) and it is expressed specifically in the developing anther compared to *NST1* (Mitsuda *et al.* 2005) which has more expression in a number of other tissues, such as interfascicular fibers of inflorescence stems, woody tissues, xylems (Mitsuda *et al*. 2007) and involvement in pod shattering in siliques (Mitsuda and Ohme-Takagi 2008).

1.4.1.2 NAC SECONDARY WALL–PROMOTING FACTOR1 (NST1) and NST2

As previously suggested, originally *NST1* and *NST2* were identified as playing a role in secondary cell wall thickening but their exact regulation was unclear (Mitsuda *et al*. 2005). However, now it appears that *MYB26* may drive *NST1/NST2* expression which via regulation of lignin and cellulose biosynthesis genes drives secondary cell wall thickening (Yang *et al.* 2007). Looking at expression levels in different tissues of *NST1/NST2* it was previously suggested that the two work to promote secondary cell wall thickening in different tissues – *NST1* promotes thickening in xylem and other non-anther tissues whereas *NST2* seems to have a more specific role in anther tissue secondary thickening (Mitsuda et al. 2007). However, although it may be that only *NST1* is required for secondary thickening to occur in xylem tissues, both *NST1* and *NST2* are able to promote secondary thickening in the endothecium of the anther redundantly – only double mutants showed a phenotypic difference to the wild

type plants. Additionally, in knockout mutants there was no significant phenotypic differences in stems and so it is possible that a third NST is responsible for controlling secondary cell wall thickening in tracheal elements (Mitsuda *et al.* 2005). It is known that *NST3* plays a role in secondary thickening of woody tissues in *Arabidopsis* (Mitsuda *et al.* 2007*),* and so could be responsible for tracheal secondary thickening along with *NST1,* but the lack of endothecium thickening in *nst1/nst2* plants suggests that *NST3* does not play a role in anther development.

NST1/NST2 are part of a large NAC-domain family, which is made up of plant-specific transcription factors (Olsen *et al*. 2005). Transcription factors act by up/down regulating other genes. It has been shown that *NST1* (Mitsuda *et al*. 2005) upregulation leads to a subsequent upregulation of a number of genes which are known to have a role in secondary thickening, including genes which are xylem-specific as well as genes known to be involved in the modification of xyloglucans (a factor in secondary cell walls). Increased *NST1* expression also led to an increase in expression of genes which are involved in secondary cell wall specific cellulose synthase production - *IRREGULAR XYLEM3* (*IRX3*) (Taylor *et al.* 1999), *IRX4* (Jones *et al*. 2001) and *IRX12* (Brown *et al*. 2005) highlighting its role in controlling a number of secondary cell wall specific factors. *NST1* is seen as a key factor in the development of secondary cell walls in vascular tissues (Mitsuda *et al.* 2007) and is expressed throughout developing plants. *NST2* is a orthologous gene to *NST1* and has been identified as playing a redundant role in anther dehiscence (Mitsuda *et al.* 2005). Knockout analysis of single gene mutations did not lead to a phenotype whilst double knock out of *nst1/nst2* leads to anther-indehiscent plants. This, coupled with the knowledge of how *MYB26* affects *NST1/NST2* expression levels, suggests that the NAC transcription factors are controlled by MYB26 protein and subsequently regulate secondary thickening in the anther redundantly by affecting expression levels of a number of other secondary cell wall biosynthesis

genes, as NAC transcription factors have been shown to do throughout the plant (McCahill and Hazen 2019).

The relationship between *NST1/NST2* and *MYB26* does not remain completely clear however. Overexpression of *NST2* or *NST1* in *myb26* plants could not recover the sterile phenotype, whilst overexpressing both *NST1* and *NST2* in a *myb26* background led to ectopic secondary thickening (Yang et al. 2017) in the epidermis layer. It appears therefore that *MYB26* plays an additional role in secondary cell wall thickening outside of simply upregulation of *NST1/NST2.* It could be that MYB26 and NST1/NST2 are required to form a dimer to initiate downstream activation of genes, it could acts as a stabiliser, or MYB26 could play a role in the removal of a repressor to *NST1/NST2*. Previous work (Mo 2017) has shown that *NST1* and *NST2* have a binding site for MYB26 protein, but did not bind MYB26 by ESMA, supporting the idea of a secondary function of MYB26 in downstream regulation. There are possibly other ways to investigate protein binding, such as DAP-seq, however that was not carried out here or previously. Alternatively, there may be additional proteins and genes involved in the network, for example *SAF1* has been identified as having a role in anther endothecium thickening (Kim *et al.* 2012) and may interact with both *MYB26* and *NST1/NST2.*

Another potential gene, *ANTHER DEHISCENCE REPRESSOR* (*ADR*) has recently been identified in endothecium thickening may interact with *NST1/NST2* (Dai *et al*. 2019). *ADR* plays a role in accumulation of Reactive Oxygen Species (ROS) and reduces expression throughout anther development. Overexpression of *ADR* leads to indehiscent anther and reduction in the expression of *NST1* and *NST2.* It is known H_2O_2 presence is involved in the control of anther endothecium thickening (Goto-Yamada *et al.* 2014) and it appears that *ADR* negatively regulates secondary thickening in the anther. The fact ectopic expression of *ADR* leads to altered expression levels of *NST1* and *NST2* would suggest it could play a role within this network. The role that *ADR* plays in

preventing the accumulation of ROS could also mean that it plays a role in the breakdown of tissues through PCD in anther development, and it could therefore be a link between the PCD of stomium and septum tissue whilst ensuring endothecium secondary thickening occurs – in simple terms as *ADR* expression is lowered, then ROS can accumulate and break down tissues, whilst at the same time, *NST1/NST2* expression is initiated.

NAC domain and homeobox HD-ZIP Class III (HD-ZIPIII) transcription factors are usually regarded as initiators of secondary cell wall thickening (Taylor-Teeples *et al.* 2015). This is part of a 3 tiered system of secondary cell wall thickening regulation, where these three layers of transcription factors bind with promoter sequences of "lower level" regulatory genes, but also directly upregulate secondary cell wall biosynthesis genes (Mangan and Alon 2003; Taylor-Teeples *et al.* 2015; Zhang *et al*. 2018b) (Figure 1.4). In *A. thaliana*, NAC family genes as generally regarded as most upstream layer of these feed-forward loops (McCahill and Hazen 2019). These genes, along with upregulating secondary biosynthesis genes, also upregulate the next layer of regulation - *MYB46* and *MYB83*. The third layer of regulation is the highly redundant biding of these MYB transcription factors to a suite of secondary cell wall biosynthesis genes (Zhong *et al*. 2010). Within anther endothecium secondary thickening, it has been hypothesised that MYB26 initiates *NST1/NST2* expression, which are then the initiators of this three layered regulation of secondary cell wall synthesis via MYB46 (Taylor-Teeples *et al*. 2015; Yang *et al.* 2017).

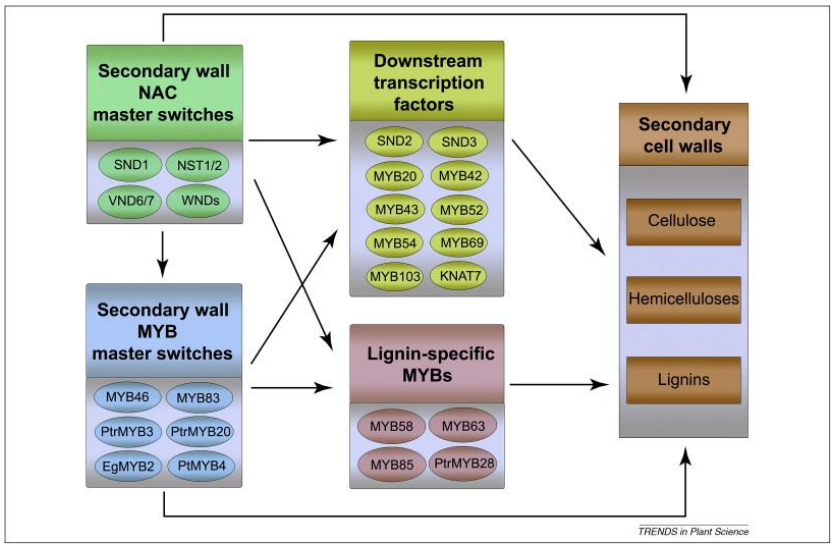

Figure 1.4 – The three layered regulation of secondary cell wall thickening in *Arabidopsis.* Figure from Zhong *et al*. (2010). This network would be downstream of *MYB26* expression, with *MYB26* activating *NST1/NST2,* in the "Secondary wall NAC master switches" section.

1.4.1.3 SECONDARY WALL THICKENING-ASSOCIATED F-BOX 1 (SAF1)

SAF1 is an F-box protein which is expressed predominantly in flower tissues. As with other F-box proteins SAF1 is responsible for protein turnover through the ubiquitin proteasome system. Through previous work it has been identified that SAF1 plays a key role in endothecium secondary thickening through negative regulation and that overexpression of SAF1 leads to anther indehiscence (Kim *et al*. 2012). This anther indehiscence is due to the down regulation of the genes required for secondary thickening – the secondary cell wall biosynthesis genes responsible for lignin or cellulose deposition – *IRX1, IRX3, IRX5, IRX6, IRX7, IRX8* and *IRX9* – genes located in the "Secondary Cell Walls" box in Figure 1.4.

1.4.1.4 Other Genes

A number of other genes are known to be involved in secondary thickening in the endothecium, generally those which are regulated by *MYB26* and

NST1/*NST2*. Genes which have expression levels altered by these transcription factors tend to therefore be genes which play some role in secondary cell wall synthesis. For example, the *IRX3*, *IRX4* and *IRX12* genes (Mitsuda *et al.* 2007; Yang *et al.* 2007) which are involved in cellulose synthase biosynthesis, a series of isoenzymes involved in secondary cell wall development (Taylor *et al*. 1999).

It is also likely that *NST1/NST2* and *MYB26* upregulate a number of other secondary cell wall genes which have been shown to be upregulated by *NST3* previously. Examples of genes which are known to be upregulated by *NST3* in vascular and woody tissues are *MYB63*, *MYB52*, and *MYB54* which have been shown to be involved in the lignin biosynthesis pathway and are therefore important in secondary cell wall deposition (Zhong *et al.* 2006). It has been suggested that *NST3* does not seem to have a role in anther endothecium secondary thickening (Mitsuda *et al.* 2005), and so it is possible that *NST1/NST2* upregulate these MYB transcription factors in this tissue. Additionally, it is likely that the suite of secondary cell wall biosynthesis genes upregulated by these MYB transcription factors, such as cellulose synthase genes, *CESA4*, *CESA7*, and *CESA8* are upregulated downstream of the three layered regulation of secondary cell wall synthesis initiated by NAC domain transcription factors (Taylor-Teeples *et al*. 2015).

Mo (2017) has also identified a number of genes which seem to interact with *MYB26.* These include a gene which have a binding motif for MYB26 which also show high expression in the anther endothecium, *PROTEIN KINASE SUPERFAMILY PROTEIN* (*PKSP*). The potentially redundant genes to *PKSP* – *PATTERN-TRIGGERED IMMUNITY (PTI) COMPROMISED RECEPTOR-LIKE CYTOPLASMIC KINASE 1* (*PCRK1*) – was also discovered. Additionally, Mo (2017) also identified proteins which were hypothesised to interact with MYB26 – namely TGACG (TGA) MOTIF-BINDING PROTEIN 9 (TGA9) (and its orthologue TGACG (TGA) MOTIF-BINDING PROTEIN 10 (TGA10)) and CHY ZINC-FINGER AND RING PROTEIN 1 (CHYR1) – based on FRET assay experiments and

expression profile analysis. These proteins seem to act with MYB26 but their roles are unclear so far.

1.4.2 Hormonal Control of Endothecium Secondary Thickening

1.4.2.1 Jasmonic Acid

As with most processes, hormones play a key role in anther and pollen development and anther dehiscence. One of the main hormones in relation to anther dehiscence is jasmonic acid (JA), as highlighted by the number of mutants involved in JA synthesis that are male sterile (Ishiguro *et al.* 2001; Park *et al*. 2002; Sanders *et al*. 2000; Stintzi 2000). One example of this is *DEFECTIVE IN ANTHER DEHISCENCE1* (*DAD1*) which encodes a lipase like protein whose expression is restricted to stamens. The *dad1* mutant is defective in both pollen maturation as well as anther dehiscence, highlighting the key role it has in anther development. *DAD1* has a role in JA biosynthesis as application of exogenous JA can recover the *dad1* mutant. It is believed that *DAD1* has a role in the extraction of linoleic acid – a JA precursor – from cellular lipids within the plant, and therefore a failure in this process leads to a failure in all subsequent processes which require JA. It is believed that once JA is synthesised in the anther that it plays a role in the late stages of anther development, including synchronising pollen release with flower opening. Despite seemingly having a role in stomium opening, it is not localised in the anther endothecium (Ishiguro *et al*. 2001) and therefore does not appear to directly influence anther endothecium thickening. This is supported by other evidence where endothecium thickening occurred as normal in the JA-defective *opr3* mutant highlighting that JA seems to involved in other aspects related to anther dehiscence, such as stomium opening (Cecchetti *et al*. 2013).

1.4.2.2 Auxin

Auxin is extremely important during many stages of anther and pollen development (Cecchetti *et al*. 2004), particularly the early stages of anther development (Cecchetti *et al*. 2008; Cheng *et al*. 2007; Nemhauser *et al.* 2000;

Okada *et al*. 1991). However, it seems that auxin mainly has an effect on anther dehiscence by negative regulation – auxin perception mutants, for example auxin response factor (*ARF6*/*ARF8*) have earlier pollen maturation and anther dehiscence along with shorter filaments than wild types (Cecchetti *et al.* 2008). Auxin transport mutants have a similar phenotype with regards to filament elongation but only a mild effect on the pollen maturation and anther dehiscence, suggesting that auxin is being produced in the anthers themselves (Cecchetti *et al*. 2008). Once auxin levels start to decline, JA levels increase, along with increased expression of *MYB26*, suggesting that auxin is repressing the expression of JA synthesis genes and *MYB26.* This could be a method of synchronising the JA dependent processes of anther dehiscence (e.g. dehydration of the epidermis and endothecium) and the MYB26-driven secondary thickening of the endothecium. Analysis by Cecchetti *et al*. (2013) found that plants with inhibited auxin response factors had early expression of *MYB26* throughout the developing anther, suggesting that the detection of auxin leads to direct repression of *MYB26,* and the removal of auxin from the system allows *MYB26* to be expressed. It seems that one major role of auxin during late stage anther development is to ensure than pollen release and maturation is synchronised with flower opening (Ishiguro *et al.* 2001). It therefore appears to play a role in endothecium thickening by delaying the expression of secondary thickening genes via repression of *MYB26* expression, and has a role in anther dehiscence by delaying JA-driven processes like stomium and flower opening, and filament elongation (Cecchetti *et al*. 2013). This is achieved by auxin from the tapetum and throughout the anther being transported to the middle layer where an auxin maxima (Cecchetti *et al*. 2017) is achieved. This, conversely, means that there may be an auxin minima in the endothecium, which can then initiate the *MYB26* network for normal secondary thickening.

1.5 Aims

Here computational models will be used to investigate whether *NST2* leads to a downregulation of *MYB26* at the transcription/translation level of regulation, or whether it acts at a post-translational level.

SAF1 will be investigated to see if, as hypothesised here, it works in the *MYB26* driven network by inhibiting the accumulation of NST1/NST2 proteins. It is also hypothesised that *MYB26* acts to inhibit *SAF1* expression, and that this is the post-translational method of allowing for the accumulation of NST1/NST2.

Downstream genes of *MYB26* along with proteins which are believed to interact with MYB26 will be investigated. It is hoped that they can be located within the currently known network and it can be deduced what role they play in anther endothecium thickening. It is hypothesised that these genes will have varying gene expression and protein accumulation levels in *SAF1, MYB26* and *NST1/NST2* mutant variations.

Chapter 2: Methods and Materials

2.1 Plant Materials and Plant Growth

2.1.1 Plant Growth Conditions

All T-DNA insertion mutants were obtained from the Nottingham Arabidopsis Stock Centre (NASC). Other plant lines (wild type or transgenic such as overexpression lines etc.) were obtained or generated by the Wilson Lab at the University of Nottingham. Any deviation from this is specified and seeds obtained elsewhere are stated in the text.

Arabidopsis thaliana seeds were sown on Levington M3 compost provided by The Scotts Company Ltd, UK, unless stated otherwise. 3 – 5 seeds were sown on 9cm plastic pots with extra plants removed after ~1 weeks' growth if too many plants were growing. Pots were covered with closed transparent plastic lid for 12-15 days as plant growth was initiated before the lid was removed and open-topped clear plastic sleeves were placed over plants to prevent cross-pollination between lines. Plants were grown in a controlled growth room with 16/8 hour light/dark cycles with a measured intensity of 180 ± 20 µmol/s/m^2 and a temperature of 23/18℃ ± 2 (day/night).

Plants containing transgenes carrying antibiotic resistance genes were selected on 1% (w/v) agar plates were made with Murashige and Skoog (MS) Basal Salt Mixture (Sigma, UK) (2.15 g salt mixture per 1 litre distilled water made up to pH 5.9 with 1M potassium hydroxide). Seeds were sterilised (section 2.1.2) before being sown on these plates. Plates were sealed with micropore tape to prevent infection before being placed in a cold room for two days to promote synchronised seed germination. Plates were then transferred to a 24 hour light growth room for ~2 weeks, before plants transfer to soil as described previously.

2.1.2 Seed Sterilisation

All steps were carried out in a laminar flow cabinet. Seeds were sterilised in Eppendorf tubes in 0.75 ml of 5% (v/v) bleach for two minutes before 0.75 ml distilled water (0.05% triton) was added and tubes were inverted 7 times. The liquid was poured out leaving a small amount with the suspended seed in the tubes. This washing with 0.75 ml distilled water (with 0.05% triton) was repeated 3-5 times to fully wash off bleach. This was checked by putting a drop of the liquid on a blue towel to ensure no discolouration. 1 ml 100% ethanol was added to tubes, shaken and poured out. 0.5 ml 100% ethanol was added to tubes before immediately being poured out onto filter paper with the seeds, which were left to dry for ~1 hour, before being scattered onto ½ MS salt plates (section 2.1.1).

2.1.3 Transformation by Floral Dipping

An overnight culture of *Agrobacterium* (5 ml in LB broth with the appropriate antibiotic) was inoculated into 100 ml of LB broth with the appropriate antibiotic, and grown in a 28℃ shaker to an optical density (OD) of 0.8-1.2 at 600 nm. 5 g of sucrose (2%) and 50 uL silwet (200 ul/L) (Lehle Seeds, US) were added and then this was used to transform floral plants using the floral dip method according to Clough and Bent (1998). After floral dipping, plants were kept away from direct light and were placed in closed sleeves for 24 hours to encourage *Agrobacterium* growth before being returned to normal growth conditions (section 2.1.1).

2.2 DNA Extraction and Genotyping

2.2.1 Crude DNA Extraction by Sucrose Buffer Method

A small (5 – 7 mm diameter), young (~3 week old) leaf was removed from *Arabidopsis* plants and place directly into 100 µl sucrose buffer (Appendix II) in PCR tubes on ice. Plant tissue was crushed using a 200 µl pipette tip and samples were then heated to 99℃ for 10 minutes. Samples were then stored

at -20℃. The supernatant could then be used as crude genomic DNA for genotyping. This DNA extraction method was adapted from Berendzen et al. (2005).

2.2.2 Genomic DNA extraction by Mini Kit

High quality genomic DNA was isolated using ISOLATE Plant DNA Mini Kits (Bioline) from 7 – 10 mm leaves from ~3 weeks old *Arabidopsis* plants. A Nanodrop ND-1000 fluoro-spectrometer (NanoDrop Technologies, USA) was used to determine the quantity and quality of DNA, which were then stored at -20℃.

2.3 Relative Gene Expression

2.3.1 RNA Extraction

Plant material (>100 mg) was collected and flash frozen in liquid nitrogen. This material was either stored in the -80℃ freezer until RNA was extracted, or RNA was immediately extracted. Total RNA was extracted using RNeasy Plant Mini Kits (Qiagen, UK) following the manufacturer's procedure except that the first use of the centrifuge was increased to a 10 minute cycle, and then the step involving the digestion of DNA was repeated twice.

To summarise, the *Arabidopsis* tissue was ground into a fine powder using an automated plastic pestle. This was performed over liquid nitrogen. The powder was homogenised in 450µl lysis buffer RLT with 1% (v/v) β-mercaptoethanol before being incubated at 56℃ for 5 minutes. The lysate was centrifuged at 13,000 x g for 10 minutes before the supernatant was combined with 500 µl 100% ethanol. This mixture was loaded into RNeasy Mini Spin Columns (provided in the kit) and centrifuged (8000 x g for 30 seconds). Columns were washed with 700 µl of buffer RW1 and centrifuged (8000 x g for 30 seconds). 10 ul RNase-Free DNase solution (QIAGEN, UK) was combined with 70 µl RDD buffer and added to columns, before being incubated for 45 minutes. This stage was repeated to ensure genomic DNA was removed. RNA bound on the columns

was washed twice (500 µl, spun down 700 µl) with buffer RPE. RNA was then eluted from the column with 20 µl RNAse-free water and stored at -80℃.

2.3.2 Complementary DNA Synthesis

RNA concentration and quality was measured using a Nanodrop ND-1000 fluoro-spectrometer (NanoDrop Technologies, USA) with 0.5 µl of RNA as for DNA quality level. RNA was mixed with 1µl oligo (dT) 12-18, 1 µl 10 mM dNTP Mix and sterile distilled water. RNA and water volumes were calculated to have a final RNA concentration of 1.5 ng RNA / µl (except in cases when RNA concentration was too low to achieve this in <11 µl, where volumes were worked out to a final concentration of 1.5 ng RNA /ul. Concentrations were kept consistent across syntheses). This mixture was incubated at 65℃ for 5 minutes before being cooled on ice for at least 1 minute. 4µl 5X First-Strand Buffer, 1µl DTT (0.1M), 1µl RNaseOUT Recombinant RNase Inhibitor and 1µl SuperScript III Reverse Transcriptase (Invitrogen, UK) were added and total mixture was incubated at 50℃ for 60 minutes, before incubation at 70℃ for 15 minutes to inactivate the reaction. cDNA samples was stored at -20℃ until required.

2.3.3 Real Time (RT) PCR

A PCR reaction mixture was made by combining either 0.1 µl cDNA (or 0.5 µl crudely extracted genomic DNA (Section 2.2.2) for genotyping) with 5 µl REDTaq 2x Master Mix (VWR, UK), 0.25 µl each primer (10pmol/µl) (usually pairs but sometimes 3 primers were used in genotyping), and molecular grade distilled water (SIGMA, UK) to a total volume of 10 µl, before being mixed by gently pipetting. PCR was carried out using a Thermos Temperature Cycler with variability occurring in annealing temperatures, extension times and cycle number (Table 2.1). Annealing temperatures were set to Tm – 3℃, extension times were set to 1 minute / kb and cycle number was 35 for genotyping and 25 for real time PCR. PCR products were analysed by gel electrophoresis.

Table 2.1 – PCR programme. Cycle number was 25 for RT-PCRs whilst 35 for genotyping PCR.

Cycle	Steps	Temperature (°C)	Duration (mm:ss)
1	Initial denaturing	94	03:00
35 / 25	Denaturation	94	00:30
	Annealing	Tm - 3	00:30
	Elongation	72	1 min / kb
1	Final Extension	72	10

2.3.4 Gel Electrophoresis

1% (w/v) molecular biology grade agarose (Melford, UK) gels in 0.5x TBE buffer (45 mM Tris-borate, 1 mM EDTA) with 0.5% (w/v) ethidium bromide (Sigma, UK) were added and cast into gel moulds. RedTaq PCR product was loaded into wells in the agarose gel along with HyperLadder™ 1kb (Bioline, UK) (to evaluate fragment size) and electrophoresed in 0.5x TBE buffer in a 100V electric field until sufficient separation of DNA fragments had occurred. Gels were placed under long wave UV light (in a Molecular Imager® Gel Doc™ XR+ System (Bioline, UK)) and DNA imaged using Image Lab™ (Bioline, UK).

2.3.5 Quantitative RT-PCR

To investigate relative gene expression quantitative PCR (reverse transcriptase qRT-PCR) using Maxima SYBR Green QRT-PCR 2x Master Mix (Thermo Fisher Scientific, UK) which includes SYBR Green I was conducted. This fluoresces when bound non-specifically to double strand DNA with increased fluorescence with increased PCR products, and so can be used to quantify the levels of amplified DNA. qRT-PCR was carried out using LightCycler® 480 System (Roche, UK). Data (.txt) was extracted and analysed on Microsoft Excel.

qRT-PCR reaction mixture consisted of 4.5 µl Maxima SYBR Green QRT-PCR 2x Master Mix, 0.2 µ l cDNA template mixed with 2 µl nuclease-free water, and 0.2 µl primer pairs (10pmol/µl) mixed with 2 µl nuclease-free water. Primer efficiency was calculated by carrying out a qRT-PCR reaction with wild type

cDNA at varying dilutions (undiluted, 1:5, 1:25, 1:125; 1:625 made using series dilutions). For relative gene expression reactions cDNA was amplified using the appropriate primer pairs in individual wells and a Light Cycler 480 multi-well plate 384 (Roche, UK); *PP2A* primers were used to amplify the *PP2A* housekeeping gene as a reference. Plates were centrifuged at 2000 g for 2 minutes and then put in the LightCycler 480 according to manufacturer's instructions. The qRT-PCR cycle is highlighted in Table 2.3. Samples were run in triplicate.

Table 2.2 - qRT-PCR programme

Cycle	Steps	Temperature (℃)	Duration (mm:ss)
1	Initial denaturing	95	10:00
55	Denaturation	95	00:30
	Annealing	62	00:30
	Elongation	72	01:00
1	Dissociation	From annealing temperature up to 95°C with a 1°C increase per cycle	00:30

2.4 Molecular Cloning Methods

2.4.1 Genomic DNA Amplification with Phusion™

For amplification of DNA to be cloned into vectors, PCR were carried out using high fidelity polymerase Phusion™ (Thermo Fisher Scientific, UK). 1 µl genomic DNA (Section 2.2.2) was combined with 2 µl HF buffer, 0.2 µl dNTPs (10 mM), 0.5 µl primer pair and 0.1 µl Phusion™ polymerase. PCR reactions were carried out using a Thermos Temperature Cycler following the conditions outlined in Table 2.3.

Table 2.3 - PCR programme for amplification with Phusion™

Cycle	Steps	Temperature (℃)	Duration (mm:ss)
1	Initial denaturing	98	03:00
35	Denaturation	98	00:10
	Annealing	Tm	00:30
	Elongation	72	30 sec / kb
1	Final Extension	72	10:00

2.4.2 TOPO® Cloning and Gateway® Technology

2.4.2.1 TOPO® Cloning

TOPO cloning is based around using the enzyme DNA topoisomerase I, which can act as both a ligase and a restriction enzyme. Following the pCR™8/GW/TOPO® TA Cloning Kit protocol (Invitrogen, UK) Entry Clones were generated with the appropriate gene (amplified DNA from Section 2.4.1). Following the manufacturer-supplied protocol for transformation of chemically competent *E. coli.* 4 µl of amplified PCR product (with a 3' adenine added) was mixed with 0.5µl salt solution, 0.5µl TOPO® vector before being incubated for 30 minutes at room temperature. 2 µl of the reaction product was then added to 100 µl DH5α *E. coli* and left on ice for 30 minutes before being heat shocked at 42℃ for 30 seconds. Transformed DH5α were added to 250 µl lysogeny broth (LB) media and incubated in a 37℃ shaker. After 1 hour this was spread onto selective 1% (w/v) LB agar plates and placed in a 7℃ incubator overnight. Individual colonies were confirmed for DNA insertion and orientation through colony PCR (Section 2.4.3) and sequencing.

2.4.2.2 Gateway® Technology

Entry vectors were isolated from transformed DH5α (Section 2.4.5) and inserts cloned into the relevant destination vectors using a Gateway BP/LR Clonase™ II Enzyme Mix (Invitrogen, UK) according to the supplied instructions to perform Gateway cloning. Generally, 2 µl entry vector, 0.5 µl destination vector were added to 1.5 µl H_2O and combined with 1 µl LR Clonase II enzyme mix. This was incubated at 25 ℃ overnight in a PCR machine before the reaction was

terminated by adding 1µl proteinase K and incubating for 10 mins at 37°C. Vectors were transformed into DH5α as TOPO® vector (Section 2.4.2.1).

2.4.3 Colony PCR

To check for the insertion of the desired vector containing the relevant DNA into DH5α or *Agrobacteria* a colony PCR was carried out. Colony PCRs were similar to genotyping PCRs (Section 2.3.3) with 5 µl REDTaq 2x Master Mix (VWR, UK) being combined with 0.25 µl each primer (10pmol/µl), and 4.5 µl H_2O. Using a 200 µl pipette tip, a small sample of individual colonies from overnight plates was mixed into this REDTaq PCR mix. A PCR reaction was carried out using a Thermos Temperature Cycler with variability occurring in annealing temperatures, extension times and cycle number (Table 2.7).

Table 2.4 – PCR programme for colony PCR with REDTaq 2x Master Mix

Cycle	Steps	Temperature (°C)	Duration (mm:ss)
1	Initial denaturing	94	04:00
35	Denaturation	94	00:30
	Annealing	52	00:30
	Elongation	72	1 min / kb
1	Final Extension	72	10

2.4.4 DNA Extraction of Plasmids and Sequencing

To extract plasmid DNA a GenElute™ Plasmid Miniprep Kit (Sigma, UK) was used following the manufacturer instructions. DNA was eluted in 50 µl elution buffer. The DNA concentration and quality was determined using a Nanodrop ND-1000 fluoro-spectrometer (NanoDrop Technologies, USA). This was diluted with water to get 10 µl of 100ng/µl DNA which was sequenced from both ends of the insert (Source Biosciences (Nottingham); analysis was carried out using SnapGene Software (GSL Biotech LLC, US).

2.4.5 Electroporation Transformation of *Agrobacterium*

40 µl *Agrobacterium* strain GV3101 (Genomics facility, University of Nottingham) was combined with 1.5 µl destination vector DNA (100 ng / µl) on

ice. This mixture was transferred into an electroporation cuvette. Using a Micropulser, an electrical pulse was run across the *Agrobacteria* cells for 5 m/s. 1 ml LB was added to the cuvette before incubation for 3 hour at 28℃. The transformed *Agrobacterium* was then plated out on selective antibiotic LB agar plates.

2.4.6 Glycerol Stock Storage of Transformed Bacteria

Glycerol stocks were generated by growing an individual DH5α or GV3101 colony in 100 ml of LB with selective antibiotics overnight. 1 ml of this culture was added to 0.5 ml of 50% (w/v) glycerol in a 1.5 ml Eppendorf tube. This was flash frozen in liquid nitrogen and stored at -80℃.

2.5 Microscopy, Dissection and Staining

2.5.1 Dissection Microscopy Use

For photographing plants during phenotyping, a Zeiss Stemi SV 6 Stereo Zoom Microscope was used with a camera attached.

2.5.2 Alexander Staining

To test for pollen viability in developed anthers Alexander staining was carried out. 20-40 µl of Alexander staining solution (Alexander 1969) was dropped onto excised anthers and analysed using a light microscope after staining for ~30 minutes. Alexander stains pollen grains which are viable pollen red/purple, whilst non-viable pollen was stained black/green or non-stained.

2.6 Development of a CRISPR Line

2.6.1 Phusion™ Reaction of pCBC-DT1T2 vector

A Phusion™ reaction was set up similar to Section 2.4.1, to enable the insertion of sgRNA into pCBC-DT1T2. Primers were designed with a 20 base pair sgRNA section flanked by DNA complementary to the vector at the 5′ end and an

overhang with site specific *BsaI* site at the 3' end of the sgRNA section. The components listed in Table 2.5 were combined together and a PCR reaction was carried out using a Thermos Temperature Cycler following the cycle outlined in Table 2.3.

2.6.2 GoldenGate Reaction

150 ng of the PCR product from Section 2.6.1 was added to the components highlighted in Table. 2.6. A GoldenGate reaction (Table 2.7) was carried out in a Thermos Temperature Cycler.

Table 2.5 – Components which were combined for DNA amplification of sgRNAs into plasmid pCBC-DT1T2

Component	Volume (µL)
Water	up to 50
5x HF buffer (1.5 mM $MgCl_2$ in final reaction)	10
dNTPs (10 mM)	1
Forward primer (10 pmol / µl)	1.25
Reverse primer (10 p mol / µl)	1.25
Template DNA	1:50 diluation of miniprepped pCBC-DT1T2
Phusion Polymerase	0.1

Table 2.6 - Components which were combined for GoldenGate reaction to clone sgRNAs into destination vector pHEE401E.

Component	Volume (µl)
Cutsmart buffer (NEB)	1.5
10 mM ATP	1.5
pHEE401E	1.5
Bsal-HF	1
T4 DNA ligase	1
Water	to 15

Table 2.7 – PCR programme for GoldenGate reaction to insert sgRNA inserts in destination vector pHEE401E.

Cycle Number	Temperature (℃)	Duration (mm:ss)
50	37	05:00
	16	05:00
1	50	05:00
1	80	05:00

2.7 AGI Locus Codes

Table 2.8 - details the AGI codes for genes used throughout this thesis

Full Name	**Common name**	**AGI Code**
MYB DOMAIN PROTEIN 26	*MYB26*	AT3G13890
NAC SECONDARY WALL THICKENING PROMOTING FACTOR1	*NST1*	AT2G46770
NAC SECONDARY WALL THICKENING PROMOTING FACTOR2	*NST2*	AT3G61910
SECONDARY WALL THICKENING-ASSOCIATED F-BOX 1	*SAF1*	AT3G62440
IRREGULAR XYLEM 1	*IRX1*	AT4G18780
SERINE/THREONINE PROTEIN PHOSPHATASE 2A	*PP2A*	AT1G69960
---	---	AT3G58960
---	---	AT3G58920
PROTEIN KINASE SUPERFAMILY PROTEIN	*PKSP*	AT5G03320
---	*PCRK1*	AT3G09830
TGACG (TGA) MOTIF-BINDING PROTEIN 9	*TGA9*	AT1G08320
TGACG (TGA) MOTIF-BINDING PROTEIN 10	*TGA10*	AT5G06839

Chapter 3: Computational Modelling of Anther Endothecium Thickening Genetic Networks

3.1 Introduction

3.1.1 Genetic Network

Anther and pollen development are key factors in male fertility in flowering plants (Scott *et al*. 2004; Smyth *et al.* 1990). The development of viable pollen is tightly regulated and coordinated with a number of physiological changes within the anther leading to the release of viable pollen at the appropriate time in floral development. One of these physiological changes is secondary cell wall thickening in the anther endothecium and a number of genes involved in this process have previously been identified.

MYB26 (Steiner-Lange *et al.* 2003; Dawson *et al.* 1993) has been identified through mutant analysis and is regarded as being the key driver in secondary endothecium thickening (Yang *et al.* 2017). *NST1* and its homologue *NST2* (Mitsuda *et al*. 2005) have also been identified to affect secondary thickening and it is suggested that *MYB26* drives expression of *NST1* and *NST2* (Yang *et al*. 2007). Further analysis has been carried out investigating the relationship between these three genes, and whilst *NST1/NST2* expression is driven by *MYB26,* it is not simply a case of MYB26 leading to transcription of *NST1/NST2.* Instead it appears that *MYB26* has to be present for the accumulation of NAC transcription factors in the cell since overexpression of NST1/NST2 fail to initiate secondary thickening in the *myb26* line, suggesting that *MYB26* plays an additional role in secondary cell wall biosynthesis (Yang *et al*. 2017). Yang *et al.* (2017) also suggested that *NST2* also promotes *MYB26* expression because overexpression of *NST2* led to an increase in *MYB26* expression (network illustrated in Figure 3.1).

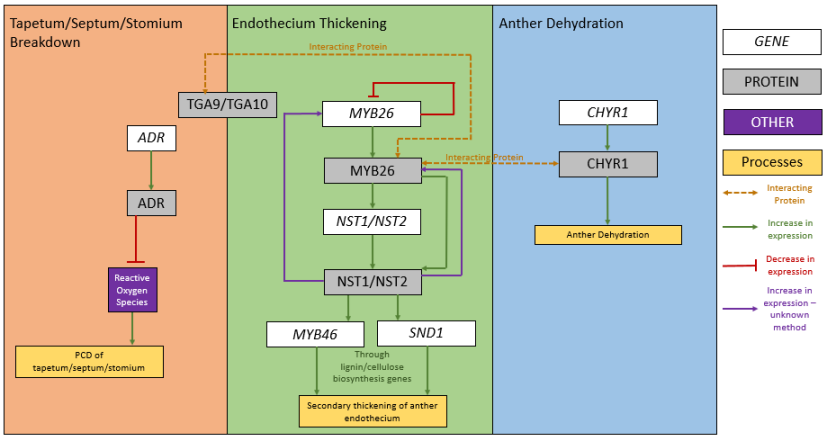

Figure 3.1 – A visualisation of the proposed network. Note the unknown upregulation of *MYB26* by *NST2* (i.e. is it increasing MYB26 accumulation at the genetic or the protein level).

Whilst Yang *et al.* (2017) investigated the relationship between *MYB26* and *NST1/NST2*, the details of how this may occur remain unclear. This can be investigated using computational models of possible scenarios involving these genes and comparing them to the observed data. In simple terms gene transcription and translation within an organism can be described by an intrinsic gene expression rate multiplied by any additional factors (such as upstream genes which upregulate/downregulate gene expression or proteins) that lead the gene to be expressed at a higher/lower level. This transcription leads to translation and protein production, which is the next stage that gene expression can be affected. Proteins also have an intrinsic rate of breakdown, and the rate of this protein accumulation are included in gene expression (for example, the theoretical situation of a gene being translated to result in one functional protein per second, whilst protein degradation is occurring at 2 per second then the translation of the gene would be negated). This can be inserted into a mathematical models to investigate how gene expression is controlled by external factors (i.e. do upstream genes change gene expression of the

downstream genes by promoting/inhibiting gene transcription or do they alter protein accumulation rate).

Here, rate of change differential equations are used to simulate gene interactions between *MYB26, NST1,* or *NST2.* These simulations can be used to generate graphical representations of relative gene expression levels of *MYB26, NST1,* and *NST2* in various mutant plants, which can then be compared to actual observed relative expression levels from previous work (Yang *et al.* 2017) to try and determine interactions between these three genes. It should be noted that at this stage the mRNA and protein expression has been considered together, and so transcription expression level and translation are equal for ease of modelling.

3.1.2 MATLAB and Modelling

By using mathematical models the transcription of genes and the translation of mRNA to proteins (which were assumed to have a direct relationship and so were grouped together for the simplicity of models) coupled with their degradation can be examined. In addition to this, by changing the equations to include other factors (e.g. expression of upstream genes) and where these factors are within the equations then the potential effects of changing gene expression of one gene on another gene can be investigated. By then comparing the theoretical models and how gene expression changes across scenarios with the observed data then the relationship between genes can be more closely understood. For example, by moving the location of parameters reliant on upstream gene expression to transcriptional or post-translational sections of simulated rate of change equations and analysing the simulated expression can suggest where gene expression of the upstream genes changed the accumulation of the relevant protein (i.e. do upstream genes affect the protein levels of the gene in question at the transcription/translation or at the post-translation stage). This can be achieved because the basic layout of these rate of change equations is:-

$$\frac{dx}{dt} = ax - bx$$

where ɑ is the rate of transcription of the gene and translation of the mRNA to the protein (which are assumed to be directly proportional for practical reasons within developing this model and are therefore grouped together) and b is the degradation of proteins (for example through ubiquitination), therefore by changing whether upstream gene is inserted into this equation in section ɑ or section b can simulate whether the upstream gene increases/decreases protein accumulation of *x* at the transcription/translation (*a*) or the post-translational level (*b*).

MATLAB is a multi-paradigm programming tool developed by MathsWorks, primarily for numerical computing, but with the addition of the Simulink package can also be used for graphical representations, and model based designs. The MATLAB programme uses a computation script which is also called MATLAB. There are a few reasons why MATLAB is a popular computational programme for network modelling, particularly the graphical outputs it can generate which makes it easier for comparison of real observed data to simulated data here.

3.1.3 Aims

In this chapter a number of mathematical models were used to simulate different gene expression combinations to investigate the topology of the network and attempt to confirm suggestions of network interactions by Yang *et al.* (2017). The models have been designed using previously observed data relating to the relationship between *MYB26*, *NST1* and *NST2*. Whilst the relationship between the promotions of *NST1/NST2* expression by *MYB26* was hypothesised by Yang *et al.* (2017), the way *NST2* promotes *MYB26* expression was not investigated with regards to transcription or post-translation upregulation in that study. These models generate theoretical relative expression graphs for different mutant plant lines, which can be compared to

the observed data generated by Yang *et al.* (2017). Comparisons of these can be used to try and confirm interaction details between genes involved in secondary anther thickening as suggested previously to identify the gene topology within this network and investigate gene topology. Here the precise interaction of NST1/NST2 and MYB26 is investigated to see if NST2 upregulates MYB26 accumulation at the transcription/translation stage, or post-translationally.

3.2 Methods

3.2.1 Uses of equations

A number of ordinary differential equations were coded into the MATLAB software. These rate of change equations were used to generate relative gene expression in different mutant lines. Different mutant lines were simulated by changing the parameters for intrinsic translation rate and for intrinsic rate of protein degradation (as an example, overexpression mutants for gene *X* would have a greater intrinsic transcription/translation (assuming they are directly proportional) rate (α) than the wild type plants. These relative expressions in various mutants were then programmed to be represented graphically for comparison with the observed data published by Yang *et al.* (2017).

All parameters in these models are set to wild type expression (translation rate, α = 1) unless otherwise stated (e.g. overexpression lines have a higher base translation rate due to the overexpression construct). This was determined by an in-built programme in Matlab which runs the equations with a range of parameter (set from 0 to 100) and determines which is the most similar to the observed data. In the scenarios modelled below for mimicking overexpression lines then α is set to 20, and for knockout lines α = 0. In lines heterozygous for *NST2,* the parameter for α was set to wild type (=1) expression because previous experiments suggest that *NST2* expression levels are similar (Yang *et*

al., 2017). Other parameter values were also determined using MATLAB's in-built parameter determination add-on as described above.

3.3 Results

3.3.1 Observed Expression Levels

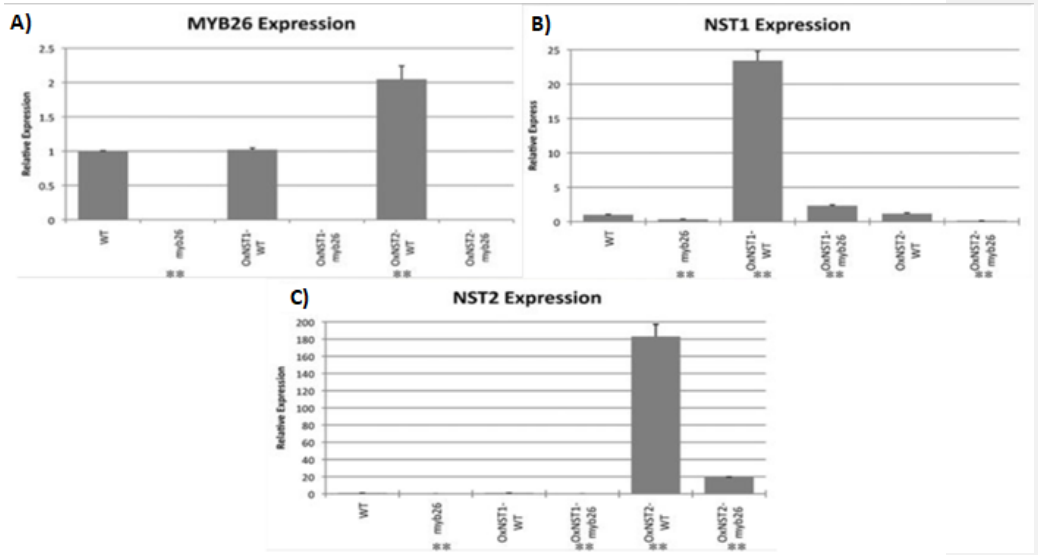

Figure 3.2 - Observed data of relative gene expression for *MYB26* (*A*), *NST1* (*B*), *and NST2* (*C*) in *Col.* wild type, myb26 knockout lines, overexpression of NST1 in wild type, overexpression of NST1 in myb26 knockout, overexpression of NST2 in wild type, and NST2 in myb26 knockout. This data is taken from Yang *et al*. (2017) Relative expression was calculated compared to a housekeeping gene (actin or PP2A). Error bars represent standard deviation (t test statistical analysis compared to its relevant background for each line; **P ≤ 0.01)

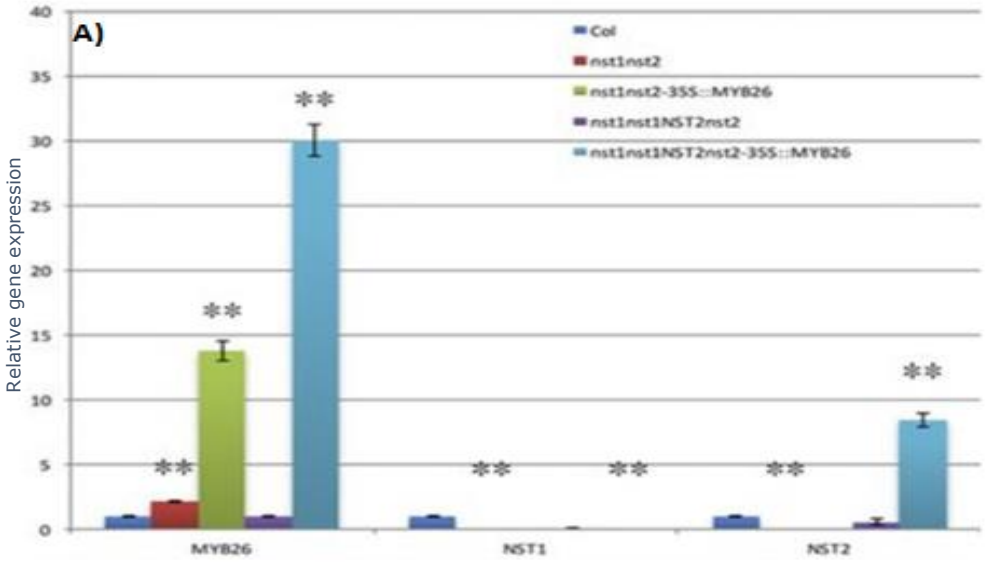

Figure 3.3 - Observed data of relative gene expression for *MYB26* in *Col*. wild type compared to an *nst1/nst2* double knockout mutant, *MYB26* overexpression in an *nst1/nst2* double knockout mutant, a *nst1nst1/NST2nst2* and overexpression of *MYB26* in *nst1nst1/NST2nst2*. It should be noted that comparison to modelled results that there was not a simulation of the *nst1nst1/NST2nst2*, but there is comparison to the other data. These graphs are taken from Yang *et al*. (2017). Relative expression was calculated compared to a housekeeping gene *PP2A.* ** denotes a $p<0.01$

Previous work (Yang *et al.* 2017) has investigated *MYB26* interactions with *NST1* and NST2 in various transgenic lines. Figure 3.1 shows observed data of relative gene expression of *MYB26* (*A*), *NST1* (*B*), and *NST2* (*C*) in wild type, *myb26* knockout lines, overexpression of *NST1* in wild type, overexpression of *NST1* in *myb26* knockout, overexpression of *NST2* in wild type, and overexpression of *NST2* in *myb26* knockout. *MYB26* and *NST2* are not expressed in the *myb26* knockout line whilst there is some *NST1* expression, though it is reduced compared to wild type. Overexpression of *NST1* in wild type plants does not lead to any change in *MYB26* or *NST2* but does, as expected, lead to a higher expression of *NST1* compared to wild type plants.

When *NST2* is overexpressed in wild type plants there is no significant change in *NST1* expression compared to the wild type, whilst *NST2* expression is increased compared to wild type. Surprisingly, an overexpression of *NST2* in wild type plants leads to an increased expression of *MYB26* compared to in wild type plants. Overexpression of either *NST1* or *NST2* in the *myb26* knockout plant lines leads to an increase in the expression of *NST1* or *NST2* respectively compared to the wild type, but the other *NAC*-domain factor gene is unaffected, along with MYB26.

Other transgenic lines (*nst1/nst2, nst1/nst2::35S::MYB26, nst1nst1NST2nst2* and *nst1nst1NST2nst2::35S::myb26*) were analysed by Yang *et al.* (2017) and the relative expression of *MYB26, NST1* and *NST2* in various transgenic lines compared to wild type plants was analysed (Figure 3.3).

MYB26 is expressed at a much higher rate in the *MYB26* overexpressed in *nst1nst1/NST2nst2* background when compared to the wild type. *MYB26* overexpression in the *nst1/nst2* double knockout leads to greater gene expression of *MYB26* compared to wild type, but significantly less than in the *nst1nst1/NST2nst2* line. Surprisingly, *MYB26* is also expressed at a higher level in the *nst1/nst2* double knockout line compared to in wild type Col. plants. *NST1* has been fully knocked out in all the mutation lines in Figure 3.3. Similarly, *NST2* is knocked out in *nst1/nst2* and *nst1/nst2* with overexpressed *MYB26.* In the *nst1nst1NST2nst2* line the *NST2* is knocked down (~0.25x) compared to the wild type but there is still some expression. When *MYB26* is overexpressed in this *nst1nst1NST2nst2* line then there is a large increase in the expression of *NST2* (~8x compared to the wild type).

3.3.2 The relationship between *MYB26, NST1,* and *NST2*

Using MATLAB, 2 different topological interactions were simulated to generate relative gene expression in a number of mutant plants, with regards to possible interactions between MYB26 and how it may control and be controlled by expression levels of NST1 and NST2. It is suggested that MYB26 drives

expression of NST1 and NST2 (Yang *et al.* 2007). It has been shown that initiating *MYB26* expression in a DEX-inducible *myb26* plant leads to an increase in *NST1/NST2* expression (after a few hours), suggesting that *MYB26* drives *NST1/NST2* expression. However, because overexpression of *NST1* or *NST2* in a *myb26* background did not rescue the indehiscence phenotype, it was concluded that MYB26 interactions with NST1/NST2 at the post-translational stage is more important. It has also been concluded that *MYB26* is self-inhibiting due to the *MYB26* expression levels in the DEX inducible line. Finally, due to the increased expression of *MYB26* when *NST2* is overexpressed in a wildtype background suggests that *NST2* may upregulate *MYB26* although it is unclear whether this is at the translation/transcription or at the post translational stage. This is investigated here. The ordinary differential equations which were used to simulate these models are stated in Figure 3.4a and Figure 3.5a whilst the genetic interactions which the equations equate to are demonstrated in Figure 3.4b and Figure 3.5b. In simple terms the models simulate interactions between MYB26 and NST1/NST2 at either the genetic, or the protein level, with:-

- Figure 3.4 - MYB26 leads to an accumulation of NST1 and NST2 by slowing down the degradation of the protein. NST2 up-regulates MYB26 by slowing down the protein degradation, and an accumulation of MYB26 slows down *MYB26* transcription.

- Figure 3.5 – As in the previous scenario, MYB26 leads to an accumulation of NST1 and NST2 by slowing down the degradation of the protein and an accumulation of MYB26 slows down *MYB26* transcription. However, in this scenario NST2 promotes an accumulation of MYB26 protein by increasing the

rate at which *MYB26* is transcribed and translated (as previously stated, these were assumed to be directly proportional for ease of model development).

a)

$$\frac{d(NST1)}{dt} = \alpha_{NST1} - \frac{K_{NST1} \cdot \beta_{NST1} \cdot NST1}{K_{NST1} + MYB26}$$

$$\frac{d(NST2)}{dt} = \alpha_{NST2} - \frac{K_{NST2} \cdot \beta_{NST2} \cdot NST2}{K_{NST2} + MYB26}$$

$$\frac{d(MYB26)}{dt} = \frac{\alpha_{MYB26} \cdot K_{M1}}{K_{M1} + MYB26} - \frac{\beta_{MYB26} \cdot K_{M2} \cdot MYB26}{K_{M2} + NST2}$$

b)

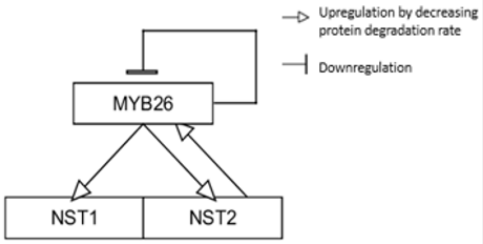

Figure 3.4 - Mathematical equations that were simulated using MatLab to model different potential interactions of *MYB26*, *NST1* and *NST2* (represented by ordinary differential equations). All equations signify the rate of change of protein accumulation over time. Equations are set out in an A= B – C layout where A is the rate of change of protein, B is the transcription and translation of the gene product grouped together and C is the removal of the protein. α is the base rate of gene translation without external input and is specific to each gene. β is the base rate of degradation of protein from within the system if there are no external factors and is also specific to each gene. K, δ, and c are constant parameters which are required to prevent equations being able to trend towards infinity and were determined using Matlab's parameter programme. Graphic representation of the above models can be found in *b* for the relationship highlighted in the equations laid out in *a*. Arrows with an open head represent an increase of protein accumulation by slowing down the removal of the target gene's protein. A bar represents downregulation of gene expression, and in all scenarios here do so by slowing down the rate of RNA production.

a)

$$\frac{d(NST1)}{dt} = \alpha_{NST1} - \frac{K_{NST1} \cdot \beta_{NST1} \cdot NST1}{K_{NST1} + MYB26}$$

$$\frac{d(NST2)}{dt} = \alpha_{NST2} - \frac{K_{NST2} \cdot \beta_{NST2} \cdot NST2}{K_{NST2} + MYB26}$$

$$\frac{d(MYB26)}{dt} = \alpha_{MYB26} \cdot \frac{c + \frac{NST2}{K_{M1}}}{1 + \frac{NST2}{K_{M1}} + \frac{MYB26}{\delta}} - \beta_{MYB26} \cdot MYB26$$

b)

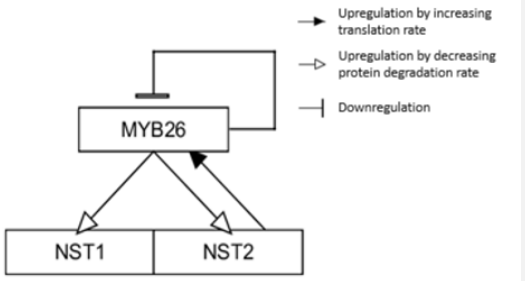

Figure 3.5 - Mathematical equations that were simulated using MatLab to model different potential interactions of *MYB26*, *NST1* and *NST2* (represented by ordinary differential equations). All equations signify the rate of change of protein accumulation over time. Equations are set out in an A= B – C layout where A is the rate of change of protein, B is the transcription and translation of the gene product grouped together and C is the removal of the protein. α is the base rate of gene translation without external input and is specific to each gene. β is the base rate of degradation of protein from within the system if there are no external factors and is also specific to each gene. K, δ, and c are constant parameters which are required to prevent equations from being able to trend towards infinity and were determined using Matlab's parameter programme. Graphic representation of the above models can be found in *b* for the relationship highlighted in the equations laid out in *a*. Arrows with solid heads represent increasing protein accumulation by increasing the rate of transcription/translation (i.e. driving gene expression) where arrows with an open head represent upregulation of gene expression by slowing down the removal of the target gene's protein. A bar represents downregulation of gene expression, and in all scenarios here do so by slowing down the rate of RNA production.

3.3.2.1 Model Variation 1

The hypothetical relative expression levels of *MYB26*, *NST1*, and *NST2* in different plant lines when a model was developed using Matlab to simulate the equations developed in Figure 3.4a to simulate the network represented in Figure 3.4b can be seen in Figure 3.6 and Figure 3.7.

Figure 3.6 shows *MYB26* is upregulated when *NST2* is overexpressed in wild type plants, but knocked out in *myb26*, even when *NST1* or *NST2* is overexpressed. This is the same as in the observed data, however the overexpression of *MYB26* is approximately twice as expressed in observed data

whereas it is approximately 15x in the model. Whilst you can change the parameters to account for this, this then leads to issues with *NST1* and *NST2* values and therefore is not an applicable solution.

Knocking out *nst1* and *nst2* in the model (Figure 3.7) leads to a knock down of *MYB26* (approximately half expression), however when a *MYB26* overexpression construct is transformed into this *nst1/nst2* line then *MYB26* expression is higher than in wild type plants (~4x compared to wild type). In the observed data knocking out *nst1* and *nst2* in the model (Figure 3.4) leads to an increase of *MYB26* (~2x compared to wild type) which is different to the model, and when *MYB26* overexpression construct is transformed into this *nst1/nst2* line then *MYB26* expression is higher than in wild type plants (as in the model), but to a greater extent (~15x compared to wild type).

When MYB26 is modelled to be overexpressed in *nst1nst1NST2nst2* line it is expressed much higher than the overexpression construct in the double homozygous knockout (*nst1/nst2*) line (~15x compared to wild type), which is also the case in the observed data, however again the observed data expresses MYB26 to a greater extent (~30x the wild type expression). Again, parameters could be used to address this but this then leads to *NST1* and *NST2* expression being less comparable with observed data.

When *myb26* is not expressed in the model *NST1* is downregulated, though it is not knocked out (Figure 3.6) with expression being approximately half that of in wild type, which is the same in the observed data. *NST1* is overexpressed (~20x) when an *NST1* overexpression construct is transformed into model wild type plants; again, this is similar to observed data (where *NST1* expression increases approximately 22 times). Whilst *NST1* has increased expression (compared to the wild type) when an *NST1* overexpression construct is transformed into *myb26* background (~10x) in the model, it is not as highly expressed as the *NST1* overexpression transformed into the wild type background(~22x). This same finding is observed in the actual collected

expression data, however with expression increases levels being different in the *NST1* overexpressed in myb26 (~2.5x) and wild type (~24x).

NST1 is similarly upregulated in the modelled expressions when *NST2* is overexpressed in the wild type background (~10x compared to wild type), but when *NST2* is overexpressed in a *myb26* background the *NST1* expression is downregulated to a similar level as in the *myb26* knockout line without any constructs (~0.5x wild type expression). The observed data shows no change in *NST1* expression compared to the wild type when *NST2* is overexpressed in the wild type background and shows reduced expression when *NST2* is overexpressed in the *myb26* background (~0.25x) which is a greater reduction than the one suggested in the model. *NST1* is completely knocked out in *nst1*/*nst2*, even when *MYB26* is overexpressed (Figure 3.9) in the model, which is the same as the expression levels seen in the collected data (Figure 3.3) based on expression analysis by Yang *et al.* (2017). Similarly, *NST1* is knocked out completely when *MYB26* is overexpressed in *nst1/NST2* in both this model tested here (Figure 3.4), and in the expression levels observed by Yang *et al*. (2017) (Figure 3.3).

In the model *NST2* expression is lower in *myb26* similar to *NST1* (~0.5x wild type expression even when *NST1* is overexpressed (Figure 3.6). This differs from the observed expression levels (Figure 3.2) where *NST2* expression is absent in *myb26*, including when *NST1* was overexpressed in this background. The model shows a big increase in *NST2* when *NST2* overexpression constructs are transformed into the wild type background (~175x wild type expression), whilst transforming *NST2* overexpression constructs into myb26 knockout plants does lead to increased expression of *NST2* but to a lesser extent than the transformed wild type background (~9x wild type expression). This is similar to the observed expression levels of *NST2*, with an increased expression of approximately 180 times higher when *NST2* is overexpressed in a wild type background, and an increase of around 20 times when *NST2* is overexpressed

in the *myb26* background reported by Yang *et al.* (2017). In the model *NST2* expression is knocked out in the *nst1/nst2* double knockout including when *MYB26* is overexpressed, but overexpression of *MYB26* in *nst1/NST2* leads to an increased expression of *NST2* (Figure 3.7) by a factor of approximately 9. The expression of *NST2* observed in the *nst1/nst2* double knockout including when *MYB26* is overexpressed shows that it is knocked out, and overexpression of *MYB26* in *nst1/NST2* leads to an increased expression of *NST2* by around 9 times – the same as in the modelled expressions.

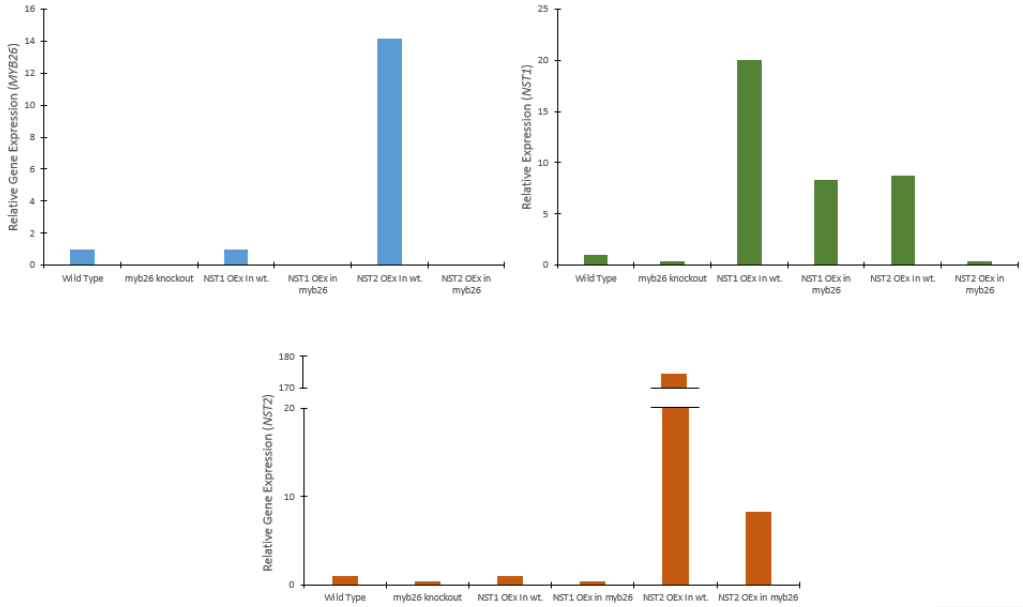

Figure 3.6 – Relative gene expression of *MYB26, NST1* and *NST2* predicted using Matlab to simulate modelling of the genetic network highlighted in Figure 3.4b using Model 1 shown in Figure 3.4a. This simulated data corresponds to the real observed data shown in Figure 3.2. For simulations base translation rate (=α) was set as 20 where genes are overexpressed, 0 for knocked out genes, and 1 for genes that were expressed to the wild type level.

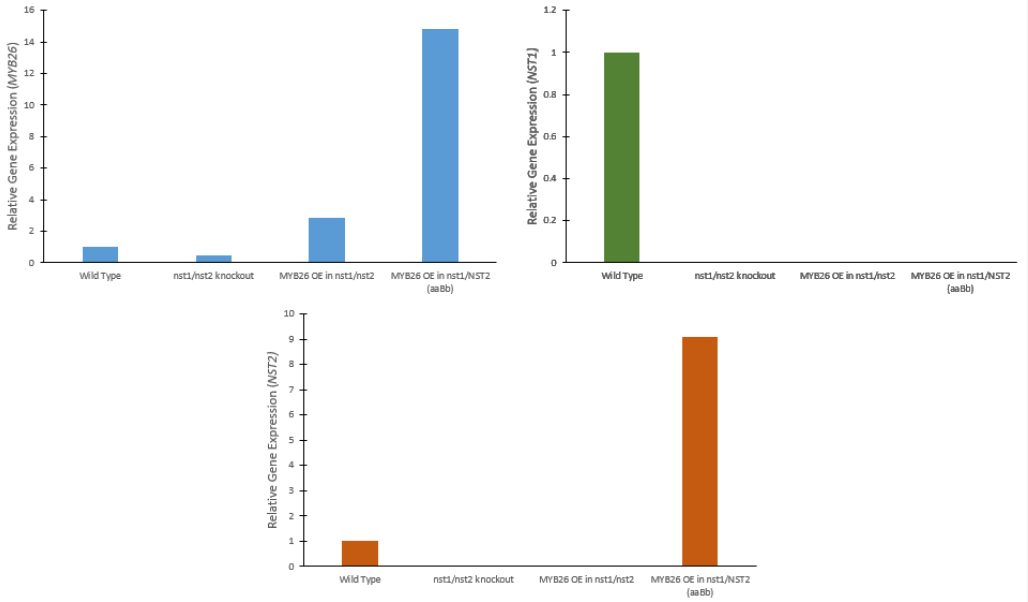

Figure 3.7 - Relative gene expression of *MYB26, NST1* and *NST2* predicted using Matlab to simulate modelling of the genetic network highlighted in Figure 3.5b using the Model 1 equations shown in Figure 3.5a. This simulated data corresponds to the real observed data shown in Figure 3.3. For simulations base translation rate (=α) was set as 20 where genes are overexpressed, 0 for knocked out genes, and 1 for genes that were expressed to the wild type level.

3.3.2.2 – Model variation 1: Discussion

Whilst modelled expression data is similar to the observed data from Yang *et al.* (2017), there are some notable differences (Table 3.1).

Commented [SF1]: Need to rewrite model variation 1 more like the second 1. Also be aware of going through and making everything past tense!!

Whilst *MYB26* expression in the modelled relative expressions tends to follow the same trend as in the observed data, modelled relative expression of *MYB26* is much higher in the model when *NST2* is overexpressed in the wild type background, whereas when *MYB26* is overexpressed in the *nst1/nst2* and *nst1nst1/NST2nst2* background the model underestimates the expression

increase. Additionally, the model predicts expression levels of *MYB26* should halve in the *nst1/nst2* plants, where in reality they actually doubled compared to the wild type.

Predicted relative expressions of *NST1* from the model are closer to the observed data, but overexpression of *NST1* in *myb26* led to a smaller increase in expression than predicted, and there was no increase in expression of *NST1* when *NST2* was overexpressed in the *myb26* background, which is different to model predictions.

NST2 is knocked out in *myb26* lines, including when *NST1* is overexpressed in this background, with expression only being knocked down, whereas the model only predicted it should be knocked out. Similarly the model predicted that overexpression of *MYB26* in the *nst1/nst2* background should lead to a knockout of *NST2*, but the observed data showed it was only knocked down. Values for increased relative expression of *NST2* in the relevant plant lines are similar to the observed data, however *NST2* was more overexpressed in reality than compared to the model.

From the comparison of the model's predicted relative expression changes and the observed data it is clear that this model is not entirely correct. The hypothesis currently is that *MYB26* is not driving *NST1/NST2* expression directly, and we know that *MYB26* exhibits self-inhibitive properties. The least understood interaction is how *NST2* drives *MYB26* expression. For this reason this interaction is changed in the second network model simulations, with *NST2* directly promoting *MYB26* expression, as illustrated in the network Figure 3.5b using the equations shown in Figure 3.5a.

Model Variation 2

Relative gene expression levels simulated by the model outlined in Figure 3.5b can be seen in Figure 3.8 and Figure 3.9. These are compared to observed data (Yang *et al*. 2017) in Table 3.2.

Within this simulation *MYB26* expression is unaffected by the overexpression of *NST1* in wild type plants, as in the actual observed data (Figure 3.2), however it is knocked down when *NST2* is overexpressed in wild type plants. This is different to the observed data where overexpressing *NST2* in a wild type background actually leads to a doubling of *MYB26* expression. In *myb26* knockout lines there is no expression of *MYB26* including when either *NST1* or *NST2* is transformed into these lines either in the simulation (Figure 3.8), or in the observed data (Figure *3.*2). In *nst1/nst2* double knockout lines *MYB26* is slightly upregulated both in the simulation (~1.5 wild type expression) and in the collected data (~2x wild type expression), and expression is greatly increased when *MYB26* overexpression constructs are transformed into *nst1/nst2* background (6x in the simulation and 14x in the observed data) and *nst1nst1/nst2NST2* background (8x in the simulation compared to 30x in the observed data) (Figure 3.9 for the simulation, Figure 3.3 for the observed data).

NST1 is downregulated in *myb26* lines (~0.5x wild type expression) in the simulation (Figure 3.7), including when *NST2* is overexpressed (~0.5x wild type expression). Whilst in the observed data (Figure 3.2) has the same expression of *NST1* in the *myb26* line as the simulation, when *NST2* is overexpressed in this background *NST1* expression actually increased by approximately 2x the wild type expression. When an *NST1* overexpression construct was transformed into *myb26* then *NST1* is expressed to a higher level than in the wild type in the simulation (x7) and the observed data (x2), however, to a less extent that when the same construct is transformed into the wild type plant (x20 in the simulation and x22 in the observed data). *NST2* overexpression in the wild type background leads to a downregulation (~0.5x wild type expression) of *NST1* in the simulation (Figure 3.8) whereas actually the observed data showed no difference*. NST1* is completely knocked out in *nst1/nst2* and *nst1nst1/NST2nst2* lines, including when *MYB26* is being overexpressed in both the simulation (Figure 3.9), and the observed data (Figure 3.3).

NST2 expression is downregulated in *myb26* in the simulation (~0.3x) whilst it is completely knocked out in the observed data (Figure 3.2), including when *NST1* is overexpressed. Transforming *myb26* with an *NST2* overexpression construct can lead to a greater expression of *NST2* than in wild type plants, although less so in the simulation compared to the observed data (~7x and 20x expression respectively). Transformed wild type background plants express *NST2* to an even greater extent (~15x in the simulation and ~180x in the collected data). *NST2* expression is knocked out in *nst1/nst2* lines in both the simulation (Figure 3.9) and the observed data (Figure 3.3), however overexpressing *MYB26* in the simulation did not recover *NST2* expression at all, whilst in the observed data the expression of *NST2* was only halved (rather than knocked out) in the *MYB26* overexpression in the *nst1/nst2* background. There is a greater expression of *NST2* compared to the wild type when *MYB26* is overexpressed in *nst1/NST2* background in both the simulation (x6) and the observed data (x9).

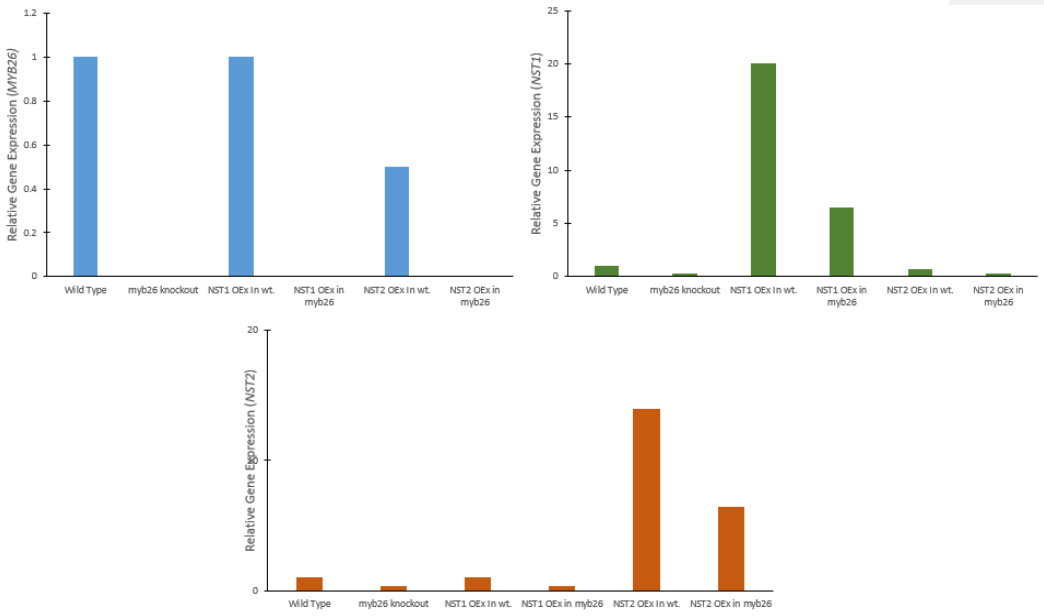

Figure 3.8 – Relative gene expression of *MYB26, NST1* and *NST2* predicted using Matlab to simulate modelling of the genetic network highlighted in Figure 3.5b using the Model 2 equations shown in Figure 3.5a. The simulated plant lines modelled here correspond to the observed data in Figure 3.2. For simulations base translation rate (=α) was set as 20 where genes are overexpressed, 0 for knocked out genes, and 1 for genes that were expressed to the wild type level.

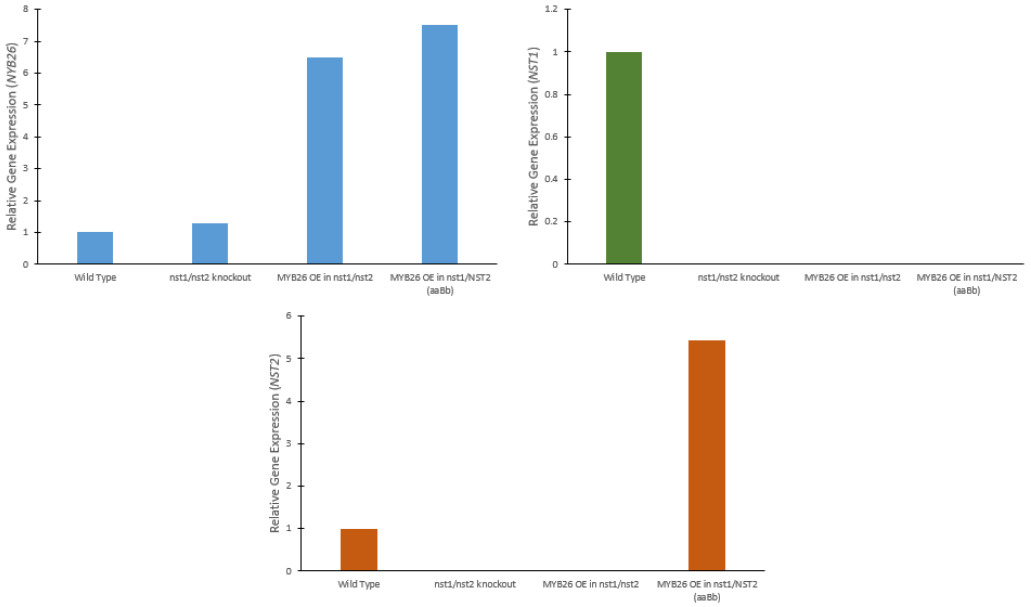

Figure 3.9 – Relative gene expression of *MYB26, NST1* and *NST2* predicted using Matlab to simulate modelling of the genetic network highlighted in Figure 3.5b using the Model 2 equations shown in Figure 3.5a. The simulated plant lines modelled here correspond to the observed data in Figure 3.3. For simulations base translation rate (=α) was set as 20 where genes are overexpressed, 0 for knocked out genes, and 1 for genes that were expressed to the wild type level.

Model Variation 2 – Discussion

In the second variation of the model the values of relative gene expression changes very similar to the data which was observed by Yang et al. (2017), however there are some differences in actual values (Table 3.2).

As with model variation 1, *MYB26* expression follows the same trend as in the observed data except for expression when *NST2* is overexpressed in a wild type background, where in the model the expression of *MYB26* is reduced (0.5x) whilst it doubles in the observed data. Other variations between the model and the observed data are fold differences with expression of *MYB26* being underestimated in the model compared to the observations in *nst1/nst2* (1.5x compared to 2x), MYB26 overexpressed in the *nst1/nst2* background (6x to

14x) and *MYB26* overexpressed in the *nst1nst1/NST2nst2* background (8x compared to 30x).

Simulated relative expressions of *NST1* from the model are closer in value to the observed data compared to Model 1. Overexpressing *NST1* or *NST2* in a wild type background leads to a slightly lower overexpression in the model (20x compared to 22x in the observations or 0.5x compared to 1x respectively), but were predicted to be higher in the *myb26* background (7x predicted compared to 2x observed in *NST1* overexpression and 0.5x predicted compared to 0.25x observed for *NST2* overexpression).

This model, like variation 1, is very similar to the observed data, particularly in terms of which gene expression levels increase/decrease in different plant lines, but there are a few differences. More notably there are issues with fold differences in the models compared to the observations. The biggest difference is generally with *NST2* overexpression. Whilst you can change parameters to increase the overexpression of *NST2* in the wild type background to get up to a similar expression as with the observed data, but this leads to further variations in other relative expression, which was detected in Matlab's automatic parameter determination add-on which is why the chosen values were used.

3.3 Discussion

Previous work carried out Yang *et al*. (2017) suggested a relationship where *MYB26* drives *NST1* and *NST2* expression at the transcriptional level, but more importantly by slowing down the removal of the NST1 and NST2 protein, although exact mechanisms are unknown. There is evidence that *NST2* in turn promotes *MYB26* expression but it was not clear if it worked at a genetic or protein level. Here 2 models were developed to investigate what the expression of some genes if *NST2* promotes *MYB26* at the post-translational (prevent protein breakdown) (Figure 3.4) or by promoting gene expression and

translation (which are grouped in this model) (Figure 3.5) level. The predicted gene expressions of *MYB26, NST1* and *NST2* seems to align more with the model variation 1 (section 3.3.1) suggesting that the *NST2* upregulation at the post-translational level. However, the actual expressions did not align perfectly with the predicted gene expressions for model variation 1 suggests that there are other factors which have to be included in the network to get a completely accurate overview.

A number of genes which may be involved in anther endothecium secondary thickening outside of the genes identified in this chapter have been identified. An F-box protein *SAF1* has been shown to lead to the same phenotype as *myb26* when it is overexpressed (Chapter 4), whilst Mo (2017) highlighted genes with interacting proteins to MYB26 along with genes which had promoter regions with a *MYB26* binding site (Chapter 6), and so it would be interesting to investigate these genes to try and incorporate them into this network, however this modelling data is useful to try and understand the hierarchy of *MYB26, NST1* and *NST2* within the network, as a new way of attempting to confirm data investigated by Yang *et al.* (2017).

3.3.1 Expanding the Equations

One concern regarding the modelled networks is the assimilation of transcription and post-translation regulation of gene expression into one equation for each gene. One thing that would be interesting to do would be to separate these out into two equations, and so an equation would be developed for mRNA and protein levels for each gene. The equations which would be used for these:-

$$\frac{d(NST1_{mRNA})}{dt} = \alpha_{NST1_{mRNA}} - \frac{K_{NST1_{mRNA}} \cdot \beta_{NST1_{mRNA}} \cdot NST1_{mRNA}}{K_{NST1_{mRNA}} + MYB26}$$

$$\frac{d(NST1_{protein})}{dt} = \sigma \cdot \varepsilon_{NST1_{mRNA}} \cdot NST1_{mRNA} - \tau \cdot \theta_{NST1_{protein}} \cdot NST1_{protein}$$

$$\frac{d(NST2_{mRNA})}{dt} = \alpha_{NST2_{mRNA}} - \frac{K_{NST2_{mRNA}} \cdot \beta_{NST2_{mRNA}} \cdot NST2_{mRNA}}{K_{NST2_{mRNA}} + MYB26_{protein}}$$

$$\frac{d(NST2_{protein})}{dt} = \lambda \cdot \varepsilon_{NST2_{mRNA}} \cdot NST2_{mRNA} - \nu \cdot \theta_{NST2_{protein}} \cdot NST2_{protein}$$

$$\frac{d(MYB26_{mRNA})}{dt} = \alpha_{MYB26_{mRNA}} \cdot \frac{c + \frac{MYB26_{mRNA}}{K_{M1}}}{1 + \frac{NST2_{protein}}{K_{M1}} + \frac{\phi \cdot MYB26_{mRNA}}{\delta}} - \beta_{MYB26_{mRNA}} \cdot MYB26_{mRNA}$$

$$\frac{d(MYB26_{protein})}{dt} = \varphi \cdot \varepsilon_{MYB26_{mRNA}} \cdot MYB26_{mRNA} - \psi \cdot \theta_{MYB26_{protein}} \cdot MYB26_{protein}$$

Above, green text within the equation are new additions to the original equations, whilst blue equations are new components. α corresponds to an intrinsic level of gene transcription, β corresponds to an intrinsic level of mRNA breakdown, ε is the intrinsic level of translation of protein from mRNA, θ is the intrinsic breakdown of protein. Other Greek letters are parameters which are used within the equations to prevent expression levels trending towards 0 or infinity. These equations could be used to predict mRNA expression levels to compare to observed data within needing to have observed protein levels, because protein levels could be calculated from the mRNA expression. If protein levels were also observed in different plant mutants these could be used to compare to predicted values, but the protein levels are still useful to calculate to provide a more accurate simulation of network with it being that by separating out these equations allows the models to be simulated more accurately and so it would be worth developing the Matlab models to simulate gene and protein interactions using these equations.

Chapter 4: Characterisation of *SAF1*

4.1 Introduction

4.1.1 *SAF1* in anther dehiscence

SAF1, a gene encoding an F-box protein, has been identified as having a role in anther secondary cell wall thickening within the endothecium (Kim *et al.* 2012) (Section 1.4.1.3). *SAF1* is predominantly expressed in floral tissues during development and is temporally confined to time between stage 13 of floral development (according to Smyth *et al*. (1990) classification), when anther dehiscence is initiating, through to stage 14. At stage 13, flowers contain anthers at phase 2 (Sanders *et al*. 1999), which include the development of endothecium thickening and anther dehiscence (Goldberg *et al*. 1993). Kim *et al*. (2012) investigated localisation of *SAF1* using a GUS reporter and found that expression was localised specifically to the endothecium layer and within pollen grains at stage 11.

To investigate the role of *SAF1* in anther endothecium thickening (Kim et al. 2012), overexpression lines of *SAF1* were generated using a cauliflower mosaic virus (CaMV) 35S promoter, resulting in male sterile plants that did not release pollen. To further confirm that this was caused by a failure of secondary cell wall thickening in the anther, they showed there was a lack of lignification in the anther endothecium of the overexpression lines compared to wild type plants at the same stage, suggesting a lack of secondary cell wall development. This hypothesis was further investigated through relative expression analysis testing for genes known to be involved in secondary cell wall biosynthesis. These were cellulose synthase genes *IRX1*, *IRX3*, and *IRX5* (Taylor *et al.* 2000; Richmond and Somerville 2000), along with glycosyl transferase genes *IRX7*, *IRX8*, and *IRX9*, and finally COBRA-like (*COBL4*/*IRX6*) genes which are all known to be involved in secondary cell wall thickening (Brown *et al.* 2005).

Knock down lines of *saf1*, although some expression was maintained, were investigated however this did not lead to a significant change in phenotype compared with the wild type, and did not affect the relative expression of the secondary cell wall thickening genes (Kim *et al.* 2012). A complete knockout was not investigated.

4.1.2 F-Box proteins and protein turnover

4.1.2.1 Protein degradation

An important aspect of regulation of gene expression in cells occurs at the post-transcriptional level - protein turnover is key for the normal functioning of cells, with proteins which have fulfilled their function needing to be removed from the cell. One particularly significant regulatory step is the removal of protein from the cell by degradation. Many of the proteins with a rapid turnover have regulatory functions, such as transcription factors (for example MYB26, NST1, NST2), whilst others are in response to a stimulus (Cooper and Hausman 2000). There are two pathways involved in the breakdown of protein from the cell, which are the lysosomal proteolysis pathway and also, particularly relevant to F-box proteins, the ubiquitin-proteasome pathway.

4.1.2.2 Ubiquitin-Proteasome System

In eukaryotes, the majority of protein degradation occurs via the ubiquitin-proteasome system (UPS) (Nandi *et al*. 2006), with the ubiquitin/26S proteasome system allowing for selective degradation of intracellular proteins (Stone and Callis 2007; Vierstra 2009). This involves proteins being degraded by a large cytosolic protease, the 26S proteasome, which is regulated by the use of a highly conserved protein ubiquitin, which tags proteins as targets for the proteasome to degrade. There are 3 stages involved in the ubiquitination of proteins, with the first stage being ubiquitin activation by an ubiquitin activation enzyme (E1). Subsequently the ubiquitin conjoined with a second enzyme, the ubiquitin-conjugating enzyme (E2). The final stage is where

specificity is conferred into the system, with a third enzyme, a ubiquitin ligase (E3) joining the ubiquitin onto the target protein (Schumann *et al*. 2011).

The ubiquitin, a 76 amino acid protein is covalently bound to the target protein with the carboxyl terminal glycine of the ubiquitin usually binding to the target protein's ε-amino group of lysine. These ubiquitin then also form similar bonds between each other with the ε-amino group of lysine of ubiquitin forming a bond with the carboxyl terminal glycine of another ubiquitin protein, leading to the formation of poly-ubiquitin chains on the target protein (Pickart 2001; Weissman 2001). Weissman (2001) also found that the lysine residue where ubiquitin molecules form poly-ubiquitin chains is an important factor in the deciding the ultimate fate of the target protein, with Lys-48 or Lys-29 linkages being signals for protein degradation. Some ubiquitin E3 ligases involved in the N-end rule pathway of targeted protein degradation in mammals (Skaar *et al*. 2009; Jin *et al.* 2004) and also, more recently in *Arabidopsis* (Zhang *et al.* 2018a*),* it has been shown that linkages of ubiquitin to the target protein can take a range of forms and have different signalling outcomes. Proteins tagged with ubiquitin for degradation are broken down by the 26S proteasome whilst the ubiquitin is released back into the cytoplasm to be recycled (Schumann *et al.* 2011).

4.1.2.3 F-box proteins in Arabidopsis

In plants the most common E3 ligases are Skp1-Cullin-F-box (SCF) protein complexes (Schumann *et al.* 2011), with F-boxes being the protein within this complex that confer specificity (Cardozo and Pagano 2004). There are a large number of F-box proteins in plants, with over 700 different F-boxes having been identified to date in *Arabidopsis* (Kim *et al.* 2012). F-box proteins have been identified as having a wide range of physiological and developmental roles in plants, ranging from plant defence mechanisms (Kim and Delaney 2002), responses to hormones (Xu et al. 2002; Dill et al. 2004; Santner and Estelle 2010), organ fusion in floral development (González-Carranza et al. 2007),

senescence and, particularly relevant to male fertility, tapetum degeneration (Kim et al. 2010) and endothecium secondary cell wall thickening (Kim et al. 2012). However, it should be noted that the majority of F-box proteins have not had their role(s) identified (Schumann *et al*. 2011).

F-box proteins are so called because the first one of them identified was human cyclin F (Elledge and Harper 1998). They share a conserved motif at their N-terminus of approximately 50 amino acids (Schumann et al. 2011), which is responsible for the recruitment of target proteins to the SCF complex, and ubiquitination. This typically leads to degradation by the 26S proteasome – for example TIR1 and AFB2 groups have been shown to upregulate auxin by causing the degradation of the Aux/IAA transcriptional repressors (Prigge *et al.* 2016). This is achieved through the regulation of Aux/IAA proteins, which have been shown to bind to transcription factors called AUXIN RESPONSE FACTORs (ARFs), which represses transcription (Salehin et al. 2015). Some F-Box proteins, for example TRANSPORT INHIBITOR RESPONSE1/AUXIN F-BOX (TIR1/AFB) within the SCF complex, binds with auxin which promotes the interaction of SCF^{TIR1} with Aux/IAA proteins, leading to their degradation and allowing the relevant transcription factors to be expressed (Kepinski and Leyser 2005). Other F-box proteins identified in *Arabidopsis* development have been shown to play a role in gibberellin (GA) signalling via *SLEEPY1* (*SLY1*) (Dill *et al*. 2004), ethylene-response pathways (Potuschak *et al.* 2003) and photoperiodic flowering through the *CONTANS* (*CO*) regulation FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of CONSTANS in *Arabidopsis* (Imaizumi *et al*. 2005).

4.1.4 Aims

Since model development in Chapter 3 could not recreate the observed relative gene expression, it is very likely that there are other factors involved within the network. *SAF1* has been previously identified as having a role in anther endothecium secondary cell wall thickening with overexpression lines having

the same phenotype as *myb26* knockout plant lines. Whilst Kim *et al*. (2012) identified *SAF1* and highlighted its role in the expression of secondary cell wall biosynthesis genes, they did not investigate the effect of *SAF1* on *MYB26* and *NST1/NST2* expression, along with any other interactions and roles it may have secondary thickening and pollen development. Additionally, *saf1* knockout lines were not investigated. In this chapter, the relationship between *SAF1* and other key drivers of anther endothecium secondary cell wall thickening is investigated with the predicted results being that SAF1 interacts in the network by removing NST1/NST2 protein, whilst itself is downregulated by *MYB26.*

4.2 Results

4.2.1 Genotyping of *saf1* insertions mutants and Overexpression of *SAF1*

A number of *saf1* insertion lines, SALK_040262, SALK _042509, SALK_038835 and SAIL_425_B06 (with insert locations highlighted in Figure 4.1) were obtained (NASC) and genotyped.

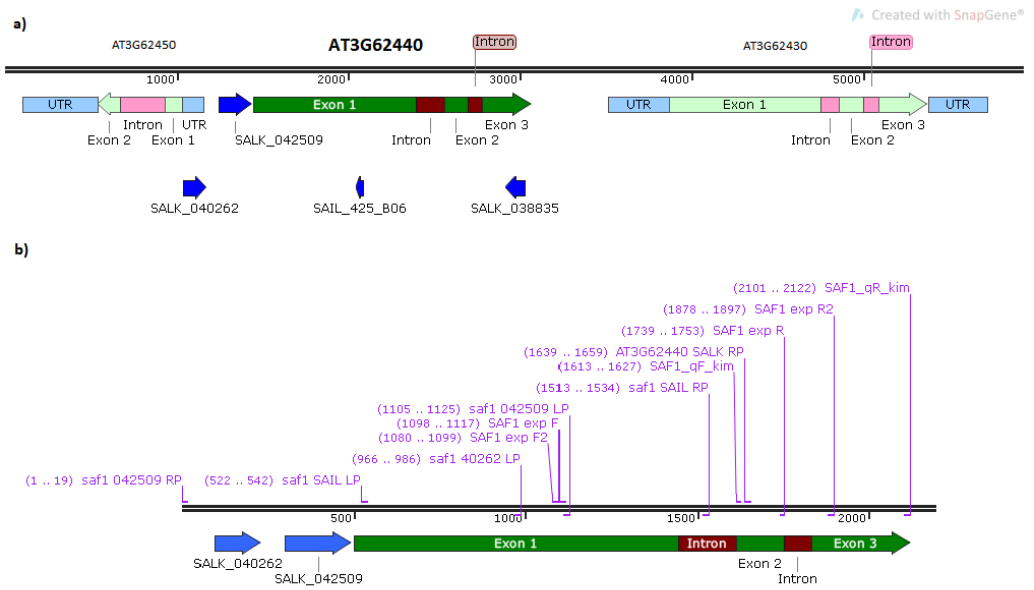

Figure 4.1 – Visualisation of the *SAF1* (*AT3G62440*) gene on chromosome 3 in *Arabidopsis thaliana*. a) highlights the flanking genes adjacent to *SAF1* b) *SAF1* and flanking sequences to show localisation of primers. T-DNA inserts are highlighted in both maps.

Eight individual plants for each SALK lines were tested with primers flanking the insert and LBb1.3 which is located in the SALK tDNA (Figure 4.1). Amplification of DNA was observed for the wild type copy of genes (986 base pair product) in both wild type and plants 2 – 7 of SALK_040262 but there was no DNA amplification associated with the SALK T-DNA insert. The same trend was observed in all SALK lines suggesting that they have all lost their T-DNA insert.

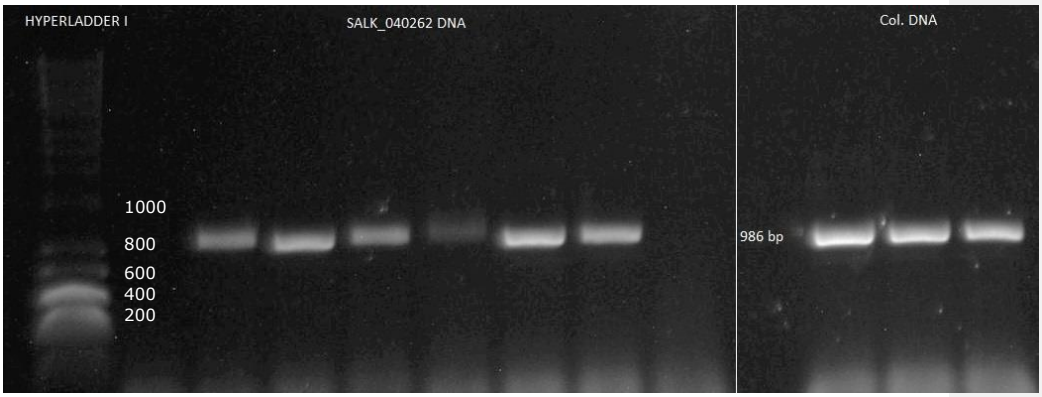

Figure 4.2 – PCR genotyping of eight individual plants of SALK_040262 compared to Col. DNA. Primers used for these amplifications were LBb1.3, saf1_042509_RP and saf1_040262_LP.

To determine if the eight individual SAIL_425_B06 plants contained the T-DNA insert highlighted in Figure 4.1a, genomic DNA was amplified (section 2.4.1.1) using *LB1* (to test for SAIL) and SAF1_SAIL_LP, SAF1_SAIL_RP primers (Figure 4.1b). Amplification of Col. wild type DNA generated a band of 1013 bp between SAF1_SAIL_LP and SAF1_SAIL_RP. The lack of a SAIL amplified insert in the wild type plants mean that there are no sequences in the DNA for the LB1 primer to bind to. In the SAIL_425_B06 plants the distance between the SAF1_SAIL_LP and SAF1_SAIL_RP primers is too long for amplification. Plants carrying a SAIL T-DNA insertion in the gene have a binding site for the LB1 primer, therefore there is amplification between LB1 and SAF1_SAIL_RP. Since it is not known exactly where within the insert the primer is located

amplification size is approximate, but it is clear that plants 1 – 4, and plants 6, 7 and 8 of the *saf1* plants contain the SAIL insert and could therefore maybe be affected in *saf1* expression (Figure 4.3).

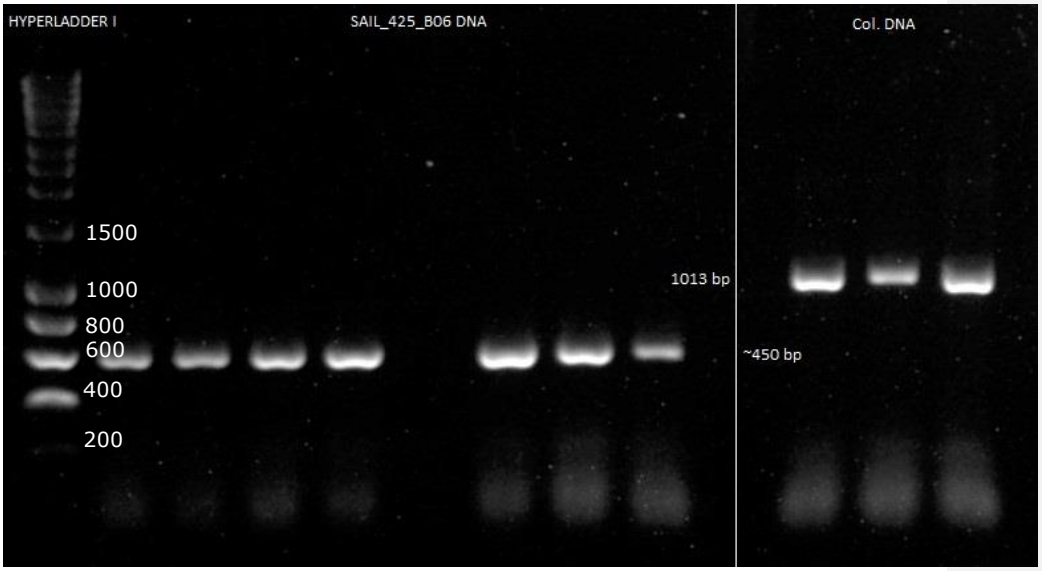

Figure 4.3 – PCR genotyping of eight individual plants of SAIL_425_B06 (*saf1*) DNA and three individual plants of Col. wild type DNA. Primers used for all samples were LB1, SAF1_SAIL_LP, SAF1_SAIL_RP (*SAF1* primers are highlighted in Figure 4.1b, LB1 is within the SAIL insert). Note that all three primers were used within the same reaction at the same time.

The SAIL lines do have the tDNA insert in plants 1 – 4, and plants 6, 7 and 8, and three of these are therefore taken forward to investigate gene expression with.

Overexpression *SAF1* plants were generated by Kim et al. (2012) using a pCAMBIA2300 vector which drives *SAF1* expression via the CaMV 35S promoter (Figure 4.4).

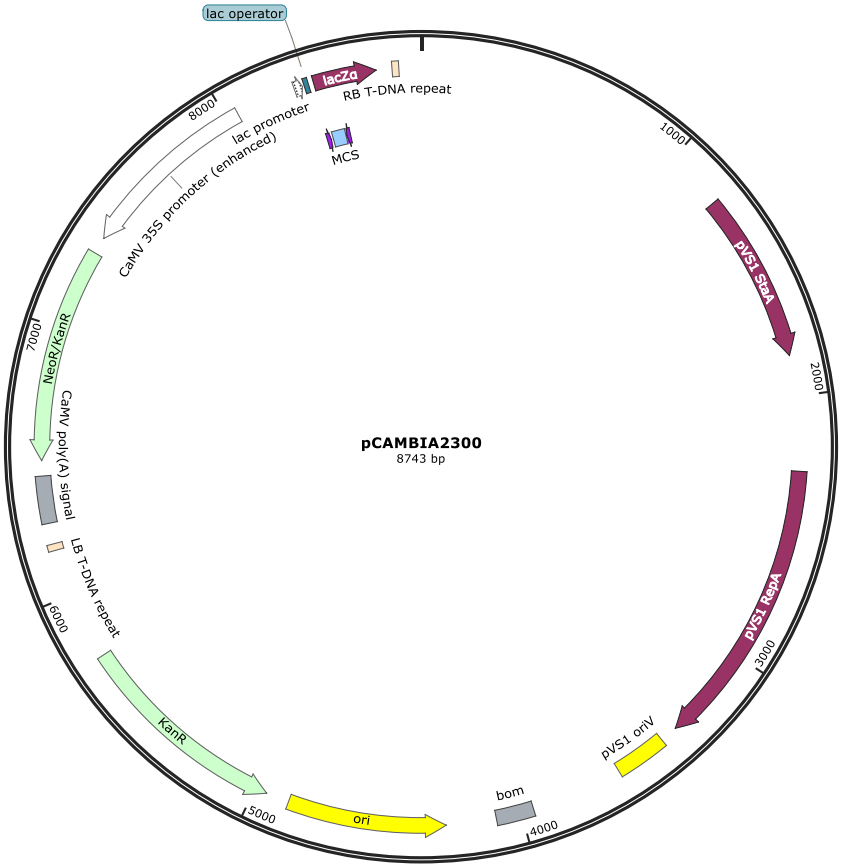

Figure 4.4 – Map of the pCAMBIA2300 vector, which was used by Kim et al. (2012) to generate *SAF1* overexpression lines. Note the 35S promoter in the vector which drives expression of these *SAF1* overexpression plants (white).

Plants were grown using seeds sent by Kim *et al.* (2012). Using a forward primer for the 35S promoter (35S_For) (located in the white section of the map (Figure 4.4)), coupled with a reverse primer for SAF1 (SAF1_40262_LP) (which is located in the *SAF1* gene), plants were tested for the presence of a 35S::SAF1 sequence. This confirmed presence of the *SAF1* construct under the control of 35S promoter (Figure 4.5). No amplification was observed in wild type plants (Figure 4.5).

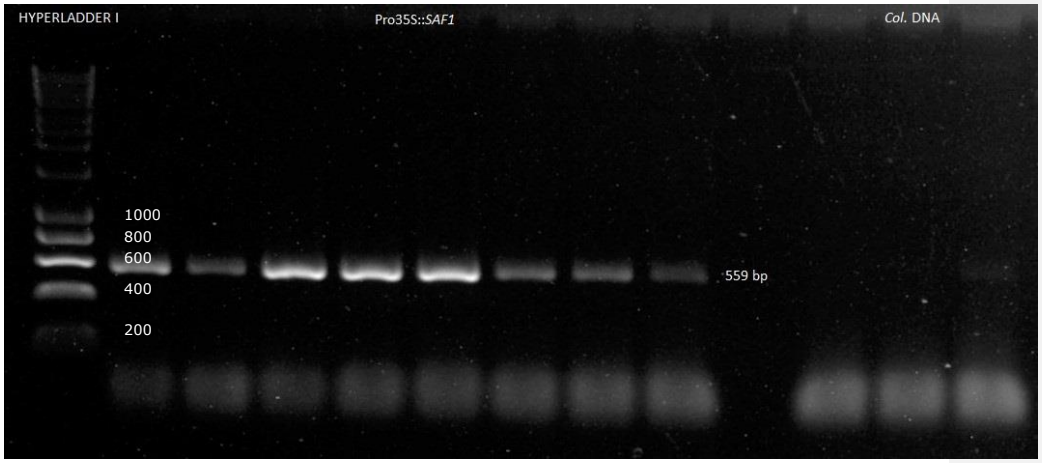

Figure 4.5 – PCR genotyping of *pro35S::SAF1* lines obtained from Kim *et al.* (2012) and Col. wild type. These were amplified using 35S_For and SAF1_40262_LP (Figure 4.1b). The expected band length is 559bp for plants with a 35S::SAF1 construct.

4.2.2 Phenotypic Analysis

One way to test for sterility in *Arabidopsis thaliana* is to observe silique lengths of developing plants. Since sterile plants will not be generating seeds within the silique they are shorter than wild type plants. When siliques of *Pro35S:SAF1* plants were measured they were very highly significantly ($p<0.001$) shorter than wild type siliques of fully grown plants, with measurements for 5 different siliques on five different biological replicates averaging 4.24mm and 13.76mm for *Pro35S:SAF1* and Col., respectively. SAIL_425_B06 (*saf1*) silique lengths (average of 13.28mm) were not statistically significantly different ($p>0.05$) to wild type (Figure 4.6). Probability values were calculated using a two-tailed Student T-Test.

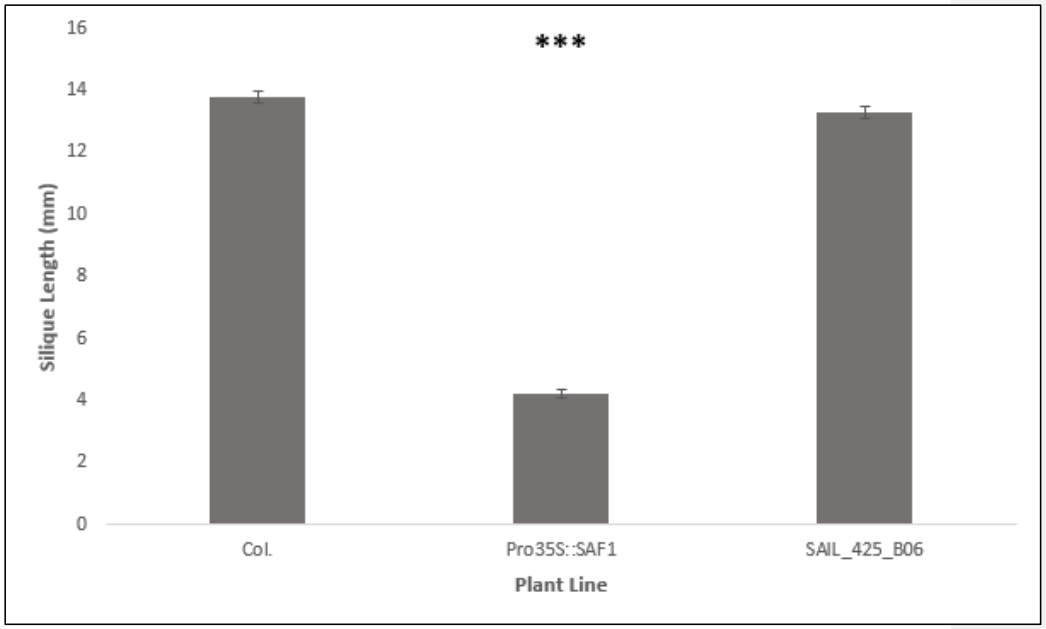

Figure 4.6 – Silique length measurements for Pro35S::SAF1, SAIL_425_B06 (*saf1*) and wild type Col. plants. There were 5 repeat measurements of 5 biological repeats from each plant line. Error bars are standard error. *** signifies the $p<0.001$ result of T-test analysis.

Closer analysis of the phenotype of the *saf1* (SAIL_425_B06) and Pro35S::SAF1 plants was carried out along with wild type plants with a specific focus on the floral development. Opened flowers from these plants were dissected (section 2.6.1) under a microscope. Wild type (Figure 4.7a and b) and *saf1* (Figure 4.9a - d) plants had pollen on the outside of anthers and on the carpel. However, flowers from the pro35S::SAF1 plants at the same stages as the other two lines, did not have pollen released from within the anther, leading to completely smooth clean anthers (Figure 8a and b). Pollen viability in the pro35S::SAF1 anthers was tested by Alexander staining (section 2.6.2); pollen was clearly visible within the Pro35S::SAF1 anther (Figure 8c), whereas in the wild type (Figure 4.7c) and *saf1* (Figure 4.9c) pollen was less closely packed within the anthers and had already been released and could be seen outside the anther. To check the viability of the pollen with the closed anthers of Pro35S::SAF1 the

anther was lightly pressed to squeeze the pollen out; the pollen stained purple was viable (Figure 4.8d). There seems to be less pollen in wild type (Figure 4.7d) and *saf1* (Figure 4.9d) anthers but this was because it was washed away during the staining since the pollen was loose on the outside of the anthers. This supports the hypothesis that overexpression of *SAF1* leads to male sterility because of a lack of pollen release, whilst knocking out *saf1* does not lead to a different phenotype compared to wild type.

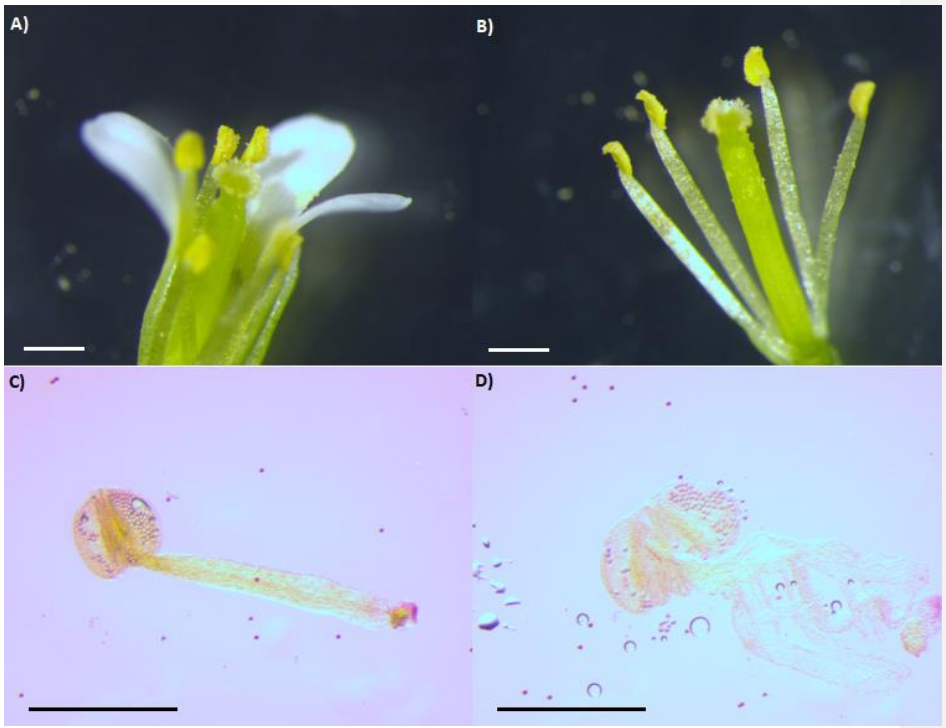

Figure 4.7 – Col. flowers dissected to view anther development. A) Front petals have been removed to allow visualisation of the stamen. B) The same flower with the rest of the petals removed. Alexander staining was carried out to test for pollen viability with C) showing an intact anther whilst D) shows the anther after it was lightly pressed to release the pollen. All pollen is stained as viable. Scale bars are 1mm.

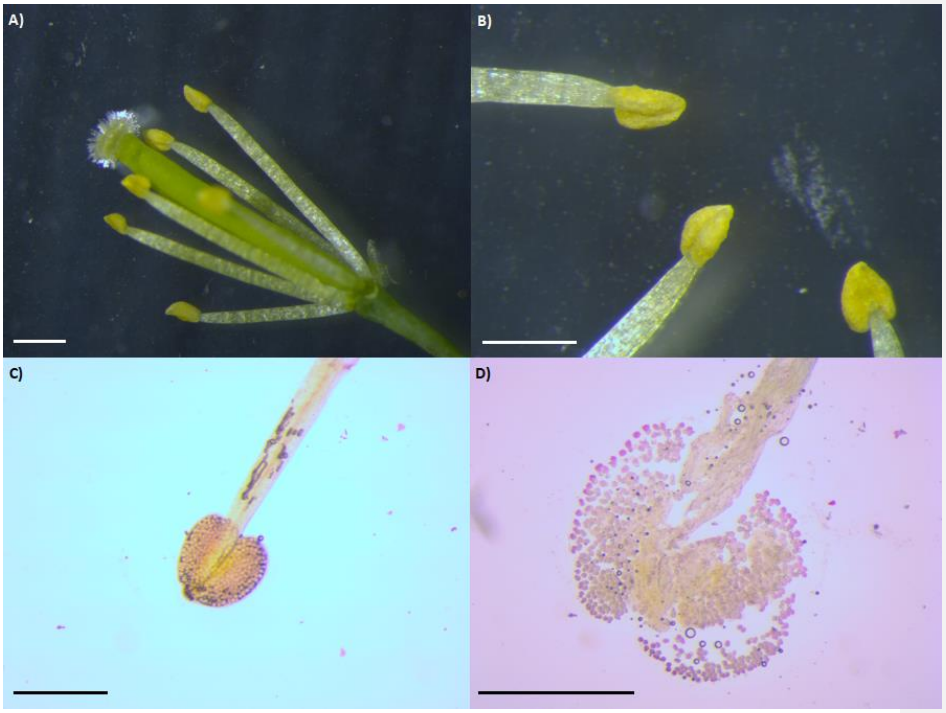

Figure 4.8 – A flower from pro35S::SAF1 which was dissected to highlight the developing anther. This shows plant number 4 from the genotyping (Figure 4.5). A) with its petals removed to allow visualisation of the stamen and the pistil. B) shows anthers from the same flower at a higher magnification to highlight the lack of pollen on the outside of the anther. Alexander staining was carried out to test for pollen viability with C) showing an intact anther whilst D) shows the anther after it was lightly pressed to release the pollen trapped inside the anther. Scale bars are 1mm.

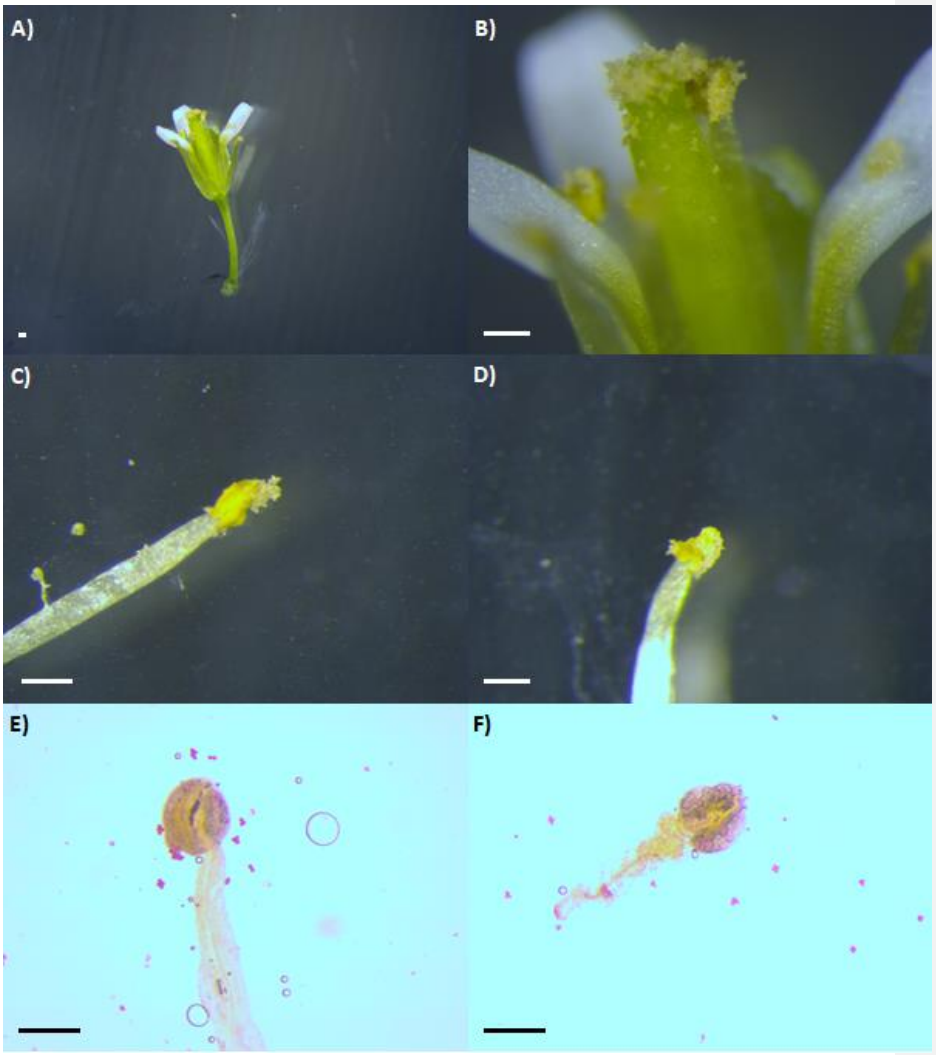

Figure 4.9 – Flowers from *saf1* plants (SAIL_425_B06). A and B) Intact Flowers C) – F) Anthers excised from the flower pictured in A) and B). A) front petals removed to allow visualisation of the stamen. B) Flower with the rest of the petals removed. Alexander staining was carried out to test for pollen viability with E) showing an intact anther whilst F) shows the anther after it was lightly pressed to release the inner pollen. Scale bars are 1mm.

4.2.3 Expression Analysis of Pro35S::SAF1 and SAIL_425_B06 plants

RNA was extracted (section 2.3.1) from pro35S::SAF1 and *saf1* (SAIL_425_B06) developing buds along with Col. wild type. cDNA was synthesized from RNA (section 2.3.2). The relative expression (compared to expression in wild type plants) of *SAF1, MYB26, NST1, NST2* and *IRX1* was examined using qRT-PCR (section 2.3.5). Inter-sample variance was normalised against the expression of the housekeeping gene *PROTEIN PHOSPHATASE 2A CATALYTIC SUBUNIT* (*PP2A*). There were 3 technical replicates and 3 different biological replicates for each gene in each plant line. Figure 4.10 shows these relative expressions with numerical values representing fold difference of genes compared to wild type. Probability values were calculated with two-tail student T-tests (Figure 4.10).

When *SAF1* was driven by the 35S promoter (in the Pro35S::SAF1) there was highly significant ($p>0.001$) overexpression (an increase of 370x) of the *SAF1* gene. When *SAF1* was overexpressed, expression of *MYB26* (0.58x expression, $p<0.05$), *NST1* (0.34x expression, $p<0.05$), *NST2* (0.26x expression, $p<0.01$) and *IRX1* (0.52x expression, $p<0.05$) were all significantly reduced.

In the *saf1* line SAIL_425_B06 there was not a significant ($p>0.05$) decrease in the expression of *SAF1* compared to the wild type. There was also no significant change in *NST1, NST2* and *IRX1* or *MYB26*. Since *SAF1* expression did not appear to be altered, a change of expression in the other genes would not be expected. There is further investigation into this lack of relative expression in the *SAF1* expression in section 5.2.1.

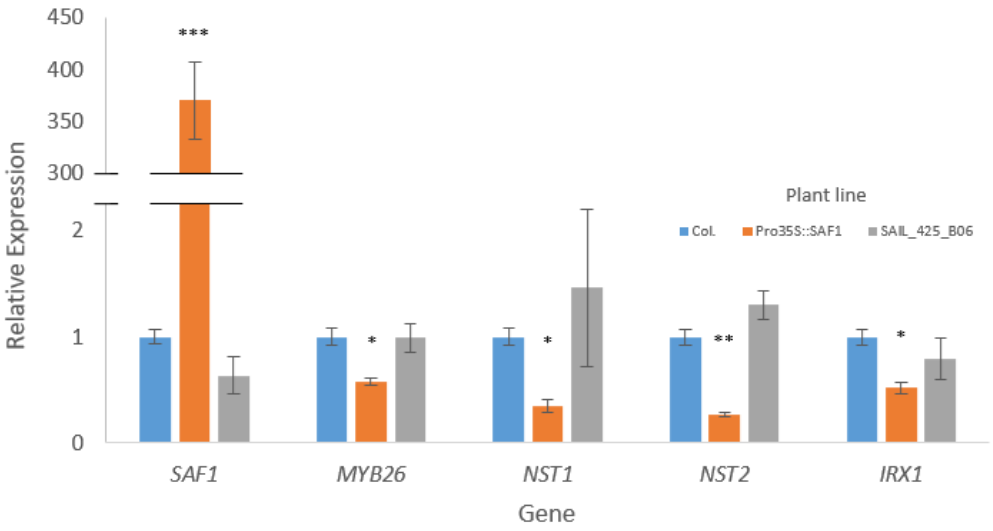

Figure 4.10 – Relative gene expression of *SAF1*, *MYB26, NST1, NST2* and the secondary cell wall biosynthesis gene *IRX1* in Pro35S::SAF1 and *saf1* (SAIL_425_B06) compared to wild type plants. There were 3 technical replicates of 3 independent plants. The housekeeping gene *PP2A* was used to normalise inter-sample variance. Error bars are standard error and probability values were calculated with two-tail student T-tests.

Expression of *SAF1* was analysed (section 2.3.5) in mutant lines of *MYB26* using a pro35S::MYB26 (generated by the Wilson lab. Previously) plant line and a *myb26* T-DNA insertion line, SALK_112372, (Figure 4.11). Overexpression of *MYB26* led to a decrease of *SAF1* (0.01x expression) and knocking out *MYB26* led to an increase in the relative expression levels of *SAF1* (19x increase) compared to in wild type plants. This supports the hypothesis that the secondary role *MYB26* plays in the accumulation of NST1/NST2 could be by preventing the removal of the proteins by the ubiquitination via the SAF1 F-box protein.

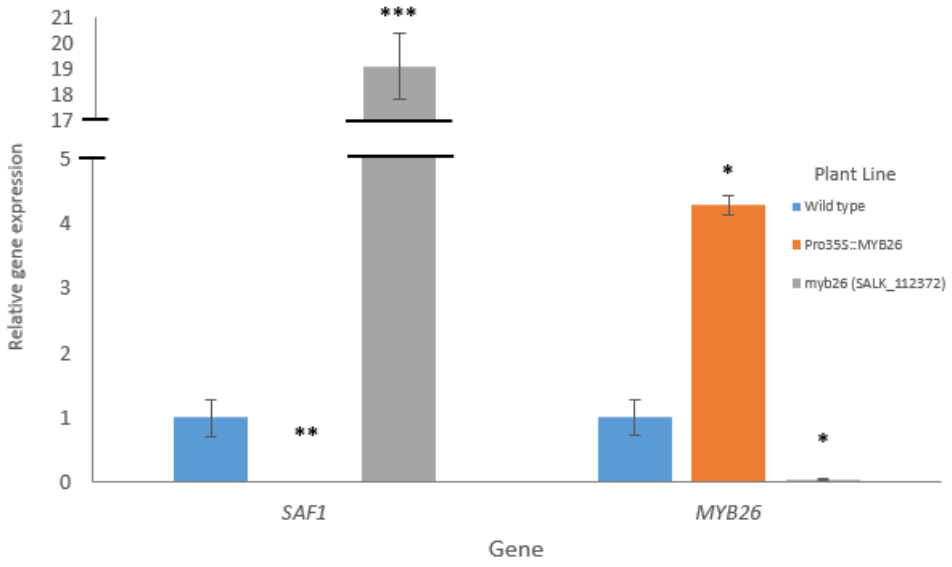

Figure 4.11 – Relative gene expression of *SAF1* and *MYB26* in Pro35S::MYB26 and *myb26* (SALK_112372). There were 3 technical repeats of 3 different individual plants for each line. The housekeeping gene PP2A was used to normalise inter-sample variance. Error bars are standard error. * represents $p<0.05$, ** represents $p<0.01$ and *** is $p<0.001$.

4.2.4 Translational Fusion of *NST1/NST2*

Translational fusions can be used to visualise protein localisation within the cell by attaching fluorescent tags to the protein of choice. Since NST1 and NST2 are important proteins in anther endothecium thickening it is very interesting to visualise how these two proteins are localised in different mutant lines. *SAF1* encodes an F-box protein and it is therefore likely to be involved in network regulation at the protein level. Based on previous work (Yang et al. 2017) and expression analysis in Figure 4.10 and Figure 4.11, this interaction is hypothesised to be achieved by SAF1 removing NST1/NST2 from developing endothecium, thereby preventing accumulation of these proteins.

Genomic DNA was extracted from wild type Col. plants (section 2.2.2) before the *NST1* or *NST2* promoter and coding sequence was amplified with high fidelity polymerase (section 2.4.1.2) using primers NST1_Pro_F and NST1_CDS_NS_R for NST1 (Figure 4.12a) and primers NST2_Pro_F2 and NST2_CDS_NS_R (Figure 4.12b) for NST2. A PCR product was successfully amplification for *NST1* (Figure 4.13a) and *NST2* (Figure 4.13b).

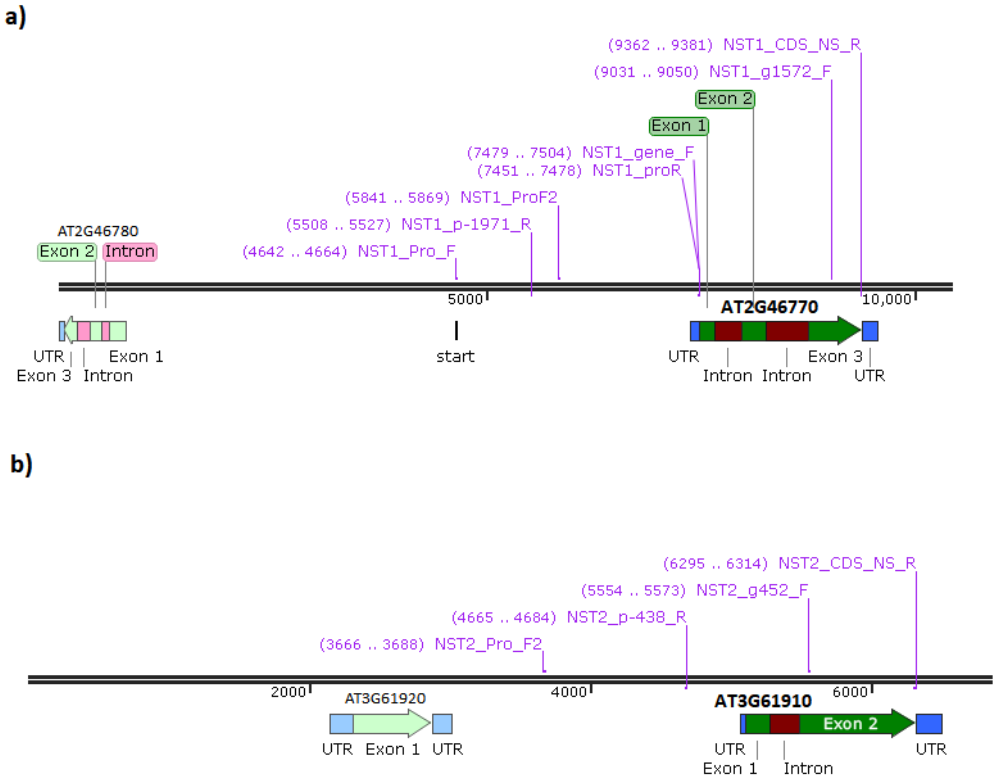

Figure 4.12 – Map of (a) *NST1* (AT2G46770) gene from chromosome 2 and (b) *NST2* (AT3G61910) gene from chromosome 3 in *Arabidopsis thaliana* with primers highlighted for both genes.

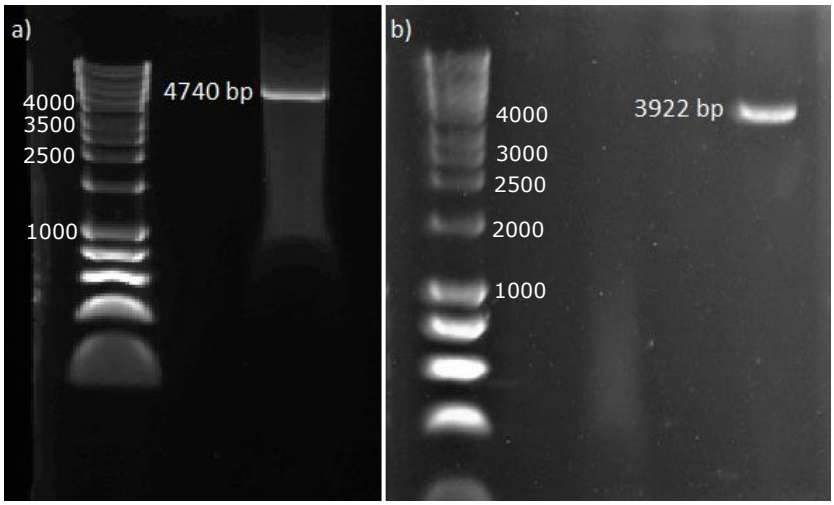

Figure 4.13 – PCR amplification of *NST1* (a) and *NST2* (b) genomic DNA. Primers used for *NST1* are NST1_Pro_F and NST1_CDS_NS_R (Figure 4.12a), whilst *NST2* was amplified using NST2_Pro_F2 and NST2_CDS_NS_R (Figure 4.12b). Ladder in both amplification checks is HyperLadder I.

NST1 (Figure 4.14a) and *NST2* (Figure 4.14b) amplified DNA was cloned into the entry vector pCR8_GW_TOPO (section 2.4.2) which was transformed into *E. coli* (section 2.4.3). Eight individual colonies were tested using colony PCR (section 2.4.4) for the presence of the appropriate insert using M13R (a primer in the M13 reverse section of the pCR8_GW_TOPO vector) and NST1_g1572_F for *NST1* (Figure 4.15a) and M13R with NST2_g452_F (Figure 4.15b). Two positive colonies (4 and 7 for *NST1*, 1 and 6 for *NST2*) were taken forwards to recombine into the appropriate destination vectors. Plasmid DNA was extracted (section 2.4.5) from these colonies and the presence of *NST1/NST2* ends confirmed by sequencing.

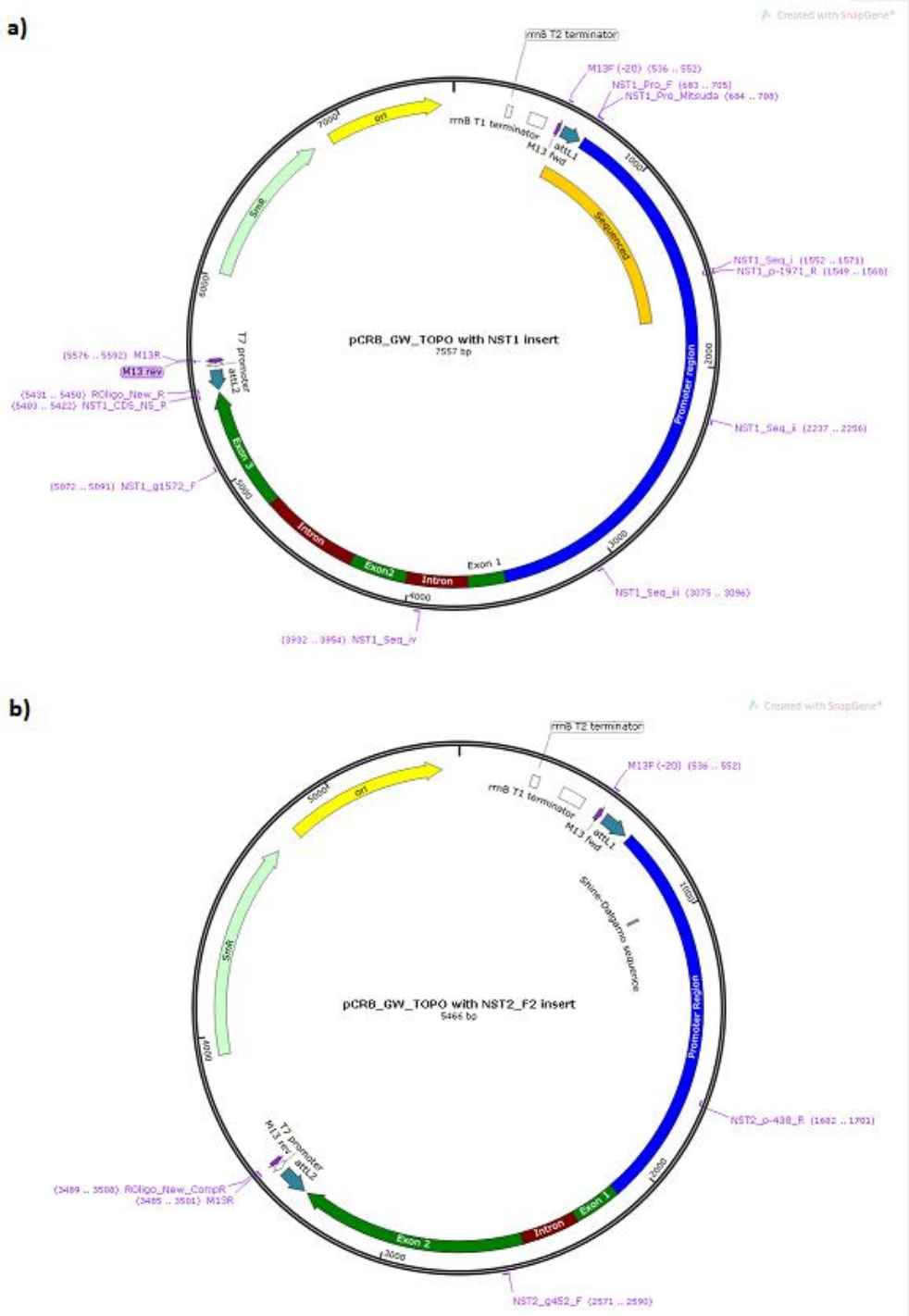

Figure 4.14 – Map of the entry vectors used in the development of translational fusion lines for (a) *NST1* and (b) *NST2.* Primers are highlighted.

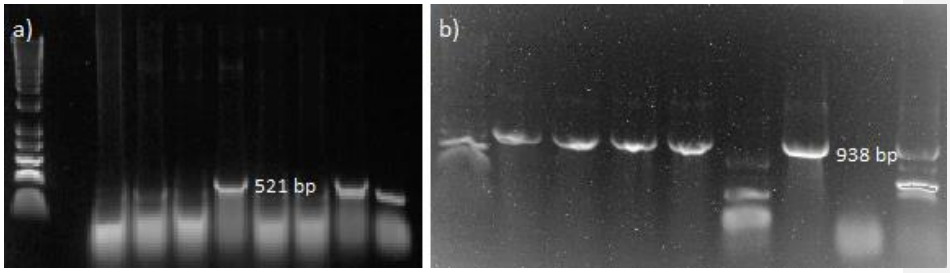

Figure 4.15 – Colony PCR amplification of pCR8_GW_TOPO vectors which had been transformed with *NST1* (a) and *NST2* (b). These were amplified using primers M13R with NST1_g1572_F for *NST1* and NST2_g452_F for *NST2*. 8 individual colonies are shown with colonies 4, 7, 8 for *NST1* and 1-4 and 6 for *NST2* showing presence of the transgene. Hyperladder I was used as the ladder in both figures.

NST1 and *NST2* sequences were recombined (section 2.4.6) into destination vectors pGHGWY and pGHGWG (Figure 4.16) respectively and transformed into *E. coli*. Colonies were grown up, plasmid DNA extracted and then inserted into DH5α *Agrobacterium*. To check for the presence of the plasmid in *Agrobacteria* PCR was carried out, and then *NST1*/*NST2* presence confirmed through sequencing. *NST2* seems to have been successfully transformed in the pGHGWG vector in *Agrobacteria. NST1* was not amplified during PCR and further sequence analysis showed that non-specific DNA sequences had been inserted into the pGHGWY vector. Sequencing from the entry vector suggested the appropriate insert was in the pCR8_GW_TOPO vector (highlighted in Figure 4.14a), however attempts to repeat this recombination reaction of *NST1* into pGHGWY were unsuccessful and could not be continued due to time limitations.

This NST2 construct was used for floral dipping of Col. and 35S::SAF1 plants (Section 2.1.3). Unfortunately when these seeds were grown on selective plates then germination occurred but seeds failed to grow past ~1 week. Even non-transformed plants should have survived for longer before selection had a notable effect, and therefore it is likely these seeds were not viable due to plant health during transformation due to unfavourable temperatures and pests in the glasshouse.

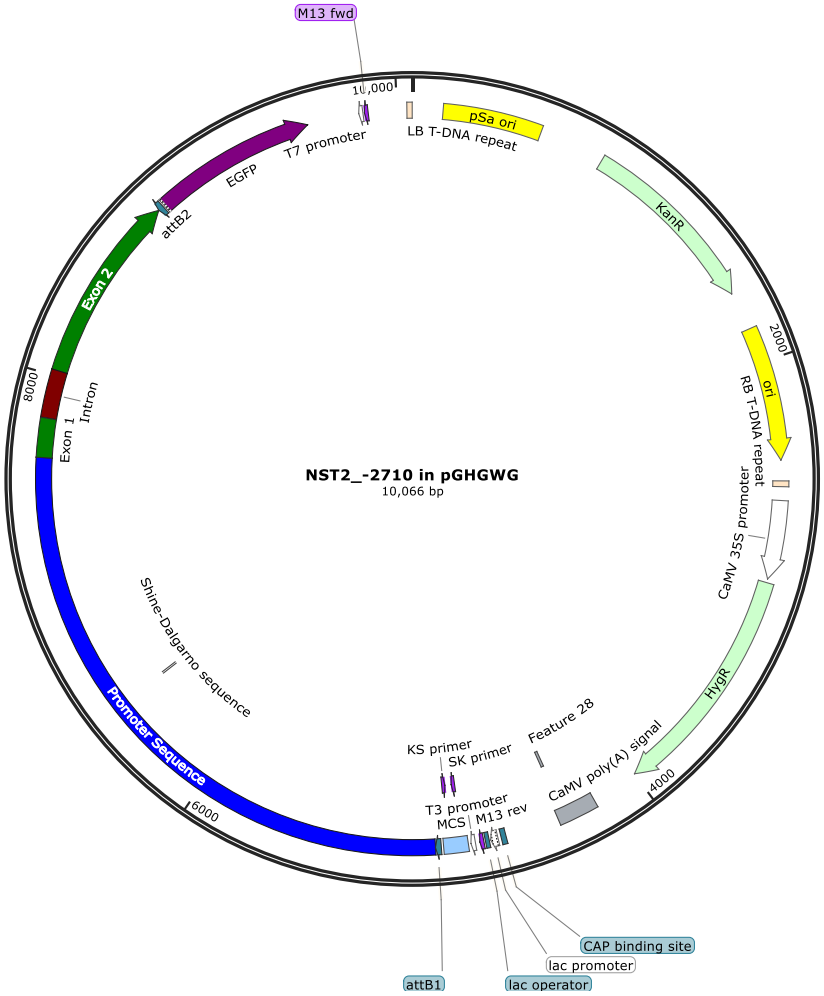

Figure 4.16 – Map of the *NST2* insert in the destination vector pGHGWG. This destination vector has a GFP protein encoded directly after the *NST2* insert, which is the reason it was used here.

4.4 Discussion

Overexpression of *SAF1* leads to plant sterility in *Arabidopsis* shown by short siliques in 35S::*SAF1* plants (Figure 4.6). This appears to be due to a lack of pollen release in the developing anther (Figure 4.8), since pollen which was extracted manually was viable when it was tested with Alexander staining and

fertility could be recovered by crossing with wild type plants or manually pollinating 35S::*SAF1* plants with pollen extracted from 35S::*SAF1* plants. In Col. wild type plants (Figure 4.7) pollen was easily released from the anther without any external influence. Previous work by Kim *et al.* (2012) had also concluded that overexpression of *SAF1* led to sterility in plants due to a lack of secondary cell wall thickening in the anther endothecium during development, which is supported by observations made here.

Kim *et al*. (2012) carried out expression analysis of how overexpression of *SAF1* changed the expression of secondary cell wall biosynthesis genes but did not investigate how expression of other genes within the network might be affected. Here, it is shown that overexpressing *SAF1* in the wild type background leads to a significant downregulation of *NST1, NST2, MYB26* and *IRX1* (Figure 4.10). *SAF1* is known to encode an F-box protein, which are involved in protein degradation within the cell (Elledge and Harper 1998) and so is likely to act at the protein level. A network of how *MYB26, NST1* and *NST2* interact with each other has been suggested by Yang *et al.* (2017) based on differences in expression in various mutant combinations which, combined with results from Chapter 3, can be seen in Figure 3.3b. As part of this suggested network MYB26 drives the expression of *NST1* and *NST2*, but NST2 also leads to upregulation of *MYB26.* The working hypothesis for where *SAF1* may interact in the network is that it acts on the NST1/NST2 proteins. This would lead to a downregulation of *MYB26* by removing the NST2 from the network, which in turn leads to all downstream genes being downregulated, as observed here.

SAF1 could also be involved in the removal of MYB26 protein from the network, however this has been rejected for now for a couple of reasons. Removal of MYB26 would lead to a reduction in the self-inhibition of *MYB26* expression (Yang *et al.* 2017), possibly recovering the expression again. Additionally, *MYB26* is known to have a secondary role in the expression of *NST1/NST2* which is why overexpression of these genes does not rescue the *myb26* phenotype.

This suggests that the downregulation of *SAF1* by *MYB26* could be the mechanism which ensures NST1/NST2 proteins can accumulate enough to drive secondary cell wall thickening, as shown by the up- and down-regulation of *SAF1* in SALK_112372 (*myb26*) and 35S::*MYB26* plants respectively (Figure 4.11).

To investigate if *SAF1* interaction at the protein level with NST2 translational fusion constructs were transformed into Col. and *35S::SAF1* plants. A construct was attempted for an NST1 translational fusion however this repeatedly failed – however, when *NST1* in pGHGWY was sequenced it was apparent that a number of foreign DNA inserts had been inserted, and thus was not as expected. However, since *NST1* is involved in secondary cell wall thickening at the plant-wide level (including vascular systems etc.) (Mitsuda et al. 2007) whilst *NST2* is anther specific and seems to be more important in secondary cell wall thickening in the endothecium (Mitsuda et al. 2005; Zhong et al. 2007; Zhong and Ye 2015), solely investigating NST2 translational fusion was therefore deemed to give a sufficient insight.

Unfortunately the transformation of 35S::*SAF1* and wild type plants with a vector containing a fluorescent tagged *NST2* protein was not successful. The vector had previously been genotyped through PCR and Sanger sequenced by Source Biosciences successfully. The seeds which had been transformed (Section 2.1.4) germinated but failed to grow past ~1 week on selection plates with hygromycin, selection by hygromycin does not usually affect seedling development until they are at least 2 weeks developed (Wilson lab, unpublished work). The lack of germination in the seeds which were transformed with the NST2::GFP vector therefore cannot be concluded to be because of a lack of successful transformation. It could be that transformed plants were not healthy and so seed germination rates were low, particularly following seed sterilisation procedures (Section 2.1.2). Nevertheless, positive transformants were not identified.

4.4.1 Future Work

The most obvious work that would be useful to carry out in the future is a repeat of the transformation of 35S::*SAF1* and wild type plants with the NST1::GFP and NST2::GFP translational fusions. It would also be useful to use this same construct to visualise the NST2 protein in wild type and other mutants, such as a *myb26* knockout line and overexpressing *MYB26* plants to confirm the positive regulation of *NST2* by *MYB26.* It would also be useful to continue to develop a similar construct to visualise NST1 protein localisation and to transform the same plants as the NST2::GFP equivalent to investigate the importance of NST1 in anther endothecium thickening compared to NST2. Alternatively, antibodies targeted to NST1/NST2 amino acid sequences along with a fluorescent tagged antibody could be used to visualise these proteins *in vivo*. One would expect that in the plants where *SAF1* is overexpressed, there would be no or reduced accumulation of NST1/NST2 protein compared to wild types, and when *saf1* was knocked out there may be increased accumulation of these proteins, although if *SAF1* expression is "turned off" by *MYB26* in the wild type, then levels may not change in the *saf1* knock out.

Another consideration in the network is the nature of the protein interactions. There are methods used to investigate these such as FRET assays or yeast-2-hybrid which can be used to investigate protein-protein interactions. Using these methods to try and confirm an interaction with SAF1 and NST1/NST2, and SAF1 and MYB26 would provide supporting evidence for the hypothesis that *MYB26* expression negatively affects the expression of *SAF1*, which in turn removes NST1/NST2 from the cell, preventing downstream expression and ultimately leading to male sterility when NST1/NST2 is not permitted to accumulate.

Chapter 5: Further Investigation of *saf1,* Potential Redundancy and Expansion of the Network

5.1 Introduction

This chapter will discuss further investigations that were carried out on the *saf1* SAIL line used in chapter 4 to try and understand why expression did not appear to be reduced. Additionally, potentially redundant genes to *SAF1* are investigated, and genes identified by Mo (2017, thesis work, Wilson Lab) are also investigated to attempt to integrate these into the network.

5.1.1 CRISPR/Cas9 edits

5.1.1.1 Gene Targeting

Arabidopsis knockout or knockdown genes are typically generated using T-DNA inserts (Krysan *et al.* 1999) which have been used in Chapter 4 in SALK and SAIL lines. However, more recently a number of genome editing technologies have emerged as ways to conduct targeted mutagenesis, for example zinc-finger nucleases (ZFNs) (Wood *et al*. 2011; Miller *et al.* 2007), and transcription activator–like effector nucleases (TALENs) (Sanjana *et al*. 2012; Wood *et al*. 2011; Boch *et al*. 2009; Christian *et al*. 2010), although due to practical difficulties these methods have not been widely adopted in the plant research community (Belhaj et al. 2013). However an alternative method of gene targeting is the RNA-guided CRISPR (clustered regularly interspaced short palindromic repeats) - Cas (CRISPR associated systems) nuclease system (Bhaya et al. 2011; Makarova et al. 2011; Horvath and Barrangou 2010), which has been gaining popularity in plant research. Whilst all three systems use double-strand breaks to interrupt genes (Ran *et al*. 2013), Cas-9 is a nuclease which can be highly specific in its action because it is guided by small RNA sequences (Jinek et al. 2012) and is being effectively applied to many species. This approach was therefore used to generate a full knock-out of the *SAF1* gene

to address issues with expression still existing in the insertional mutant lines (Section 4.2.3).

5.1.1.2 CRISPR/Cas9

CRISPR is an immune response in bacteria and archea which can be used to acquire resistance to viral and plasmid pathogens at a nucleic acid level (Sorek et al. 2013). This is achieved by integrating short fragments of pathogenic DNA into the repetitive DNA element (the CRISPR) at the end of the host chromosome. These DNA repeats are highly variable but are all short (20 – 50 base pairs) repeating sequences with a similar length of unique DNA between sequences. Whilst there is high variability between the sequences repeated between CRISPRs, there is conservation of the repeat sequence within a specific CRISPR (Kunin et al. 2007) and most CRISPR sequences contain a conserved 3'-end motif of GAAA(C/G) (Sorek et al. 2013), which may serve as a binding site for Cas proteins. Within the genome there is often an AT-rich sequence of DNA flanking the CRISPR loci which is called a "leader". These leader sequences are polar, contain promoter regions along with binding sites for regulatory proteins (Pul et al. 2010; Hale et al. 2012) which are responsible for CRISPR RNA expression and acquisition of new sequences with regards to immunity (Yosef *et al*. 2012).

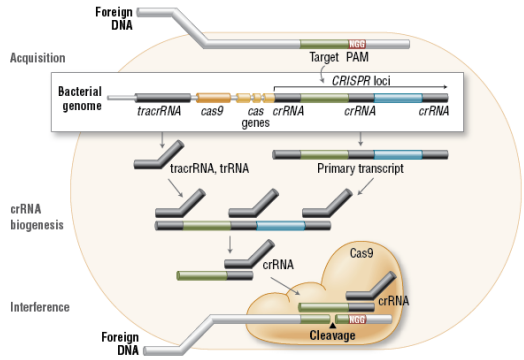

Figure 5.1 – A representation of the development of targeted DNA cleaving by Cas9 nuclease. Foreign DNA is replicated within the CRISPR loci to guide Cas9 to cleave the double strands of the invading DNA. PAM is the protospacer adjacent motif figure adapted from Reis *et al*. (2014).

Alongside the CRISPRs which have been identified, a number of CRISPR associated systems (*cas*) genes have

also been detected (Jansen *et al.* 2002), which have been categorised into three types. Particularly relevant here are the Type II CRISPR-Associated Systems, which is the best characterised (Garneau *et al*. 2010) and appear to be specific to bacteria (Makarova *et al*. 2011). There are only 4 identified *CAS* genes, one of which is *CAS9* (Sorek *et al.* 2013). The encoded CAS9 protein is a nuclease which contains guiding sequences complimentary to RNA sequences of foreign DNA, which then target the CAS9 to cleave the invading DNA at proto-spacer sites (Figure 5.1) (Reis *et al*. 2014). More recently the two RNA (crRNA and transcrRNA) motifs required to target foreign DNA have been combined to create a synthetic RNA chimera (single guide RNA, or sgRNA) which can be used to guide Cas9 (Jinek *et al*. 2012). By controlling the sequence of sgRNAs followed by a "Protospacer Adjacent Motif" (PAM) site, cleavage of specific genes by Cas9 can be achieved across species, ranging from mammals (Chen *et al.* 2011) to plants (Belhaj *et al.* 2013). When developing systems for eukaryotic cells there are a number of optimised Cas9s which can be used for different organisms. Here the maize codon-optimised pHEE401e CRISPR/Cas9 vector was used, as it has previously been shown to work successfully in CRISPR knock out gene editing in *Arabidopsis* (Wang et al. 2015).

Once DNA has been cleaved by the Cas9 nuclease it typically undergoes DNA repair via one of two pathways which can be exploited for genome targeting. The high-fidelity homologous recombination (HR) pathway can be utilised to develop precise, targeted locus modifications during repair to make small edits in the genome, such as single-nucleotide mutations (Yin *et al*. 2014). The HR pathways tends to only be active in rapidly dividing cells (Saleh-Gohari and Helleday 2004) and it is therefore especially useful in gene modification during embryo or germ line development *(*Wang *et al.* 2013). Genome editing using the HR pathway is difficult in higher plant species because of the low efficiency of homologous recombination (Fauser *et al*. 2012), although efficiency is higher

in double strand breaks as used in the Cas9 system compared to single strand breaks (Miki et al. 2017).

The alternative pathway for genome repair after cleaving by Cas9 is the non-homologous end joining (NHEJ) pathway, which is prone to errors (Ran *et al*. 2013). This error prone DNA repair can be used to establish knockout lines for genes by inserting insertion or deletion errors into DNA cleaved by Cas9, leading to frame shift mutations and premature stop codons within the gene (Perez et al. 2008; Shen et al. 2017).

5.1.2 *PROTEIN KINASE SUPERFAMILY PROTEIN* (*PKSP*)

Previous work (Mo 2017) has used co-expression analysis to identify genes which were potentially downstream of *MYB26.* One particularly gene *PKSP,* which encodes a protein kinase superfamily protein, was found to be potentially co-expressed with *MYB26*; expression of *PKSP* was down regulated in *myb26/ms35* plants and upregulated in overexpression lines of *MYB26*, suggesting it could act downstream of *MYB26* (Mo 2017). Additionally, Chromatin Immunoprecipitation (ChIP)-PCR analysis identified a MYB26-binding site in the first intron of *PKSP,* suggesting direct binding may occur and that MYB26 may therefore directly regulate the expression of *PKSP* (Mo 2017). However it is not clear whether *PKSP* is involved in the secondary cell wall thickening of developing *Arabidopsis* anthers, or whether it plays a role in a distinct network which also involves *MYB26.*

The *Arabidopsis* Genome Initiative (2000) identified over 1000 protein kinases, with estimates that 1 to 2% of functional genes in eukaryotes are encoding protein kinases, highlighting their importance (Lehti-Shiu and Shiu 2012; Zulawski et al. 2014). Protein kinases are enzymes which are responsible for removing the γ-phosphate in ATP molecules and attaching it to amino acid side chains on proteins. This impacts on how proteins interact with each other, with phosphorylated proteins interacting differently to their non-phosphorylated versions. Generally, there are two types of protein kinase in eukaryotes, some

which phosphorylate serine and/or threonine, and those that phosphorylate tyrosine in the amino acid side chains (Stone and Walker 1995).

PKSP specifically encodes a receptor-like cytoplasmic kinase (RLCK) and is predicted to phosphorylate serine/threonine amino acids. RLCKs are part of a larger superfamily of receptor-like kinases (RLKs) which are transmembrane proteins which detect signals outside the cell using receptors, which in turn lead to changes within the cell due to an intracellular kinase domain (Shiu and Bleecker 2001). The difference between RLCKs and RLKs is that RLCKs are based solely within the cytoplasm of the cell and lack the transmembrane domain present in other RLKs, however both lead to physiological changes or regulate gene expression (Lin *et al*. 2013).

Generally, RLCKs are reported to play a role in a range of aspects of plant signalling, plant immunity and stress relief, however there is an over-representation of RLCK subfamily VII, of which PKSP is one (Shiu and Bleecker 2001), in plant responses to biotic stresses (Lehti-Shiu *et al*. 2009). This could suggest that whilst it is regulated by *MYB26*, it may have a different role to that associated with anther endothecium secondary cell wall thickening. However, some RLCKs have been shown to be involved in plant growth and development (Vij *et al*. 2008) and so it may have a role in floral development.

In addition to *PKSP* (also termed *PCRK2*), a similar receptor-like cytoplasmic kinase, *PCRK1* has been as having a high level of homology to *PKSP* (Mo 2017). *PCRK1* is shown to play a role in pathogen defence, particularly against *Pseudomonas syringae* pv. *maculicola* ES4326. In *pcrk1* mutant lines there was reduced callose deposition and immunity demonstrated (Sreekanta *et al*. 2015), but this reduction did not significantly increase in *pcrk1/pcrk2* double mutants suggesting a lack of redundancy. However, Kong et al. (2016) described a significant reduction in salicylic acid (SA) in the double mutant knock out compared to the single mutants, supporting the hypothesis there is redundancy at least in some aspects of their role between these two genes.

5.1.3 MYB26 Putative Interacting Proteins

Investigation into potential interacting proteins with MYB26 were carried out using yeast-2-hybrid analysis (C. Yang. Z.A. Wilson unpublished data) and identified 5 possible interacting proteins. One of these proteins was localised to the plasma membrane and therefore unlikely to interact with MYB26 directly, whilst the other 4 possible interacting genes (*Y2H320*, *Y2H560*, *Y2H620* and *Y2H970*) were localised to the nucleus as seen with MYB26. Förster resonance energy transfer (FRET) assays were carried out on these four proteins and two (Y2H320 and Y2H560) were shown to interact with MYB26 (Mo 2017).

Y2H320 (*TGA9*) has been shown to be involved in anther development, working redundantly with *TGA10* (Murmu *et al*. 2010). Mutant analysis showed that *tga9* and *tga10* plants did not have a phenotype, whilst double knock out *tga9/tga10* lines were anther indehiscent (Mo 2017). There were abnormalities in tapetal development in *tga9/tga10* lines, along with a less turgid anther endothecium, suggesting a potential role in anther endothecium thickening. *TGA9* promoter activity has previously been observed from stage 4 – 11 of floral development, with expression peaking at stage 5 before declining through stages 7 – 11. *TGA10* has been shown to be similarly localised temporally, but to a lower level than *TGA9* (Murmu *et al*. 2010). However, analysis from Mo (2017) using qRT-PCR suggested that *TGA10* expression was actually later than described by Murmu et al. (2010). Since *TGA9* and *TGA10* work redundantly and both interact with ROXY proteins during anther development regulation it may be that a lack of *TGA9* expression could lead to a wider spread of *TGA10* expression, and *TGA10* may therefore have a potential interaction with *MYB26.* Previously *MYB26* promoter activity was shown to be detected in throughout the anther at stage 10 (Yang et al. 2007), suggesting that its expression is temporally aligned with *TGA9* expression.

Mo (2017) used light microscopy observing transverse sections of anthers to analyse *tga9/tga10* T-DNA knock out plants, noting that abnormal development

of the anther was observed from stage 5 onwards, with normal microspore mother cells produced only in the abaxial lobes. Adaxial anther lobes showed arrested development at this stage. Pollen was produced in the abaxial lobes of the *tga9/tga10* mutant, however anthers were indehiscent and therefore plants were sterile. *tga9/tga10* plants did have endothecium thickening but, as mentioned above endothecial tissue was less turgid and may be due to a reduction in secondary cell wall thickening, but cell wall composition was not analysed by Mo (2017).

Y2H560 (At5g25560) is a CHY-type/CTCHY-type/RING-type Zinc finger protein. *Y2H560* is expressed to a greater level than *MYB26* throughout anther development with a small peak in expression at the bicellular (BC) stage meaning that there is temporal and localisation overlap with *MYB26* expression (Mo 2017). The gene *CHY ZINC-FINGER AND RING PROTEIN 1* (*CHYR1*) is phylogenetically close to *Y2H560*, and also has expression during PMI stage meaning it also is temporally synchronised with *MYB26* expression. Y2H560 has no known function and knocking out this gene with a T-DNA insert does not lead to a phenotype (Wilson Lab, unpublished data). *CHYR1* positively regulates stomata opening and is involved in drought stress responses (Hsu *et al*. 2014), and so it is possible that *Y2H560* has a similar role redundantly. Since it is known that water transport is involved in late stage anther development with regards to dehiscence (Wilson *et al.* 2011) it could be possible that *MYB26* regulates this through *CHYR1* and *Y2H560* separate to anther endothecium secondary cell wall thickening.

5.1.4 Aims

Since homozygous tDNA inserts in the *saf1* line (SAIL_425_B06) didn't significantly (T-test: $p>0.05$) knock down *SAF1* expression (section 4.2.3), it is understandable that there is no difference to the wild type phenotype (section 4.2.2). Here the cDNA which was detected in qPCR experiments (Figure 4.10) is investigated to see if it is likely to be translated into a functional protein.

Genes with a possible redundancy to *SAF1* were identified and investigated with a view to bringing these into the network. Finally, to overcome the cDNA synthesised from *saf1* RNA possibly leading encoding a functional protein, gene editing by CRISPR/Cas9 was used to develop a new *saf1* mutant.

TGA9, TGA10 and *PKSP* were also examined to try to determine how they may fit into the network with *MYB26, NST1/NST2* and *SAF1* and any potentially redundant genes.

5.2 Results

5.2.1 Investigation of *saf1* SAIL_425_B06

Since the *saf1* SAIL plants did not knockout, or even significantly knock down the expression of *SAF1,* despite having the SAIL tDNA in plants at the expected location, the level and characteristics of the expression were investigated.

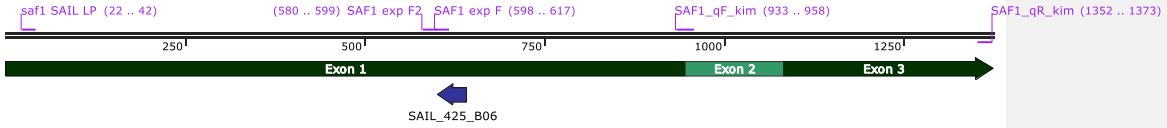

Figure 5.2 – Visualisation of *SAF1* complementary DNA with various primers highlighted, along with the location of the SAIL insert. The primer SAF1_qR_kim was used as the reverse primer with the other four primers as forward primers.

SAF1 expression was detected by qRT-PCR expression analysis in Section 4.2.3. RT-PCR of *saf1* cDNA did not generate any amplification using whole gene spanning primers (saf1_SAIL_LP and SAF1_qR_Kim) (Figure 5.2), whilst the same primers generated a band from wild type plants (Figure 5.3a). To check the *SAF1* expression in *saf1* plants, a PCR was carried out using the specific qRT-PCR (SAF1_qF_kim and SAF1_qR_kim) primers (Figure 5.2) which were used in the relative gene expression work done in Chapter 4, which resulted in DNA product amplification (Figure 5.3b). The lack of whole gene amplification, whilst expression was detected with qRT-PCR primers suggest there may be

partial transcript of *SAF1* in the *saf1* SAIL knockout plants which was detected in relative gene expression analysis (Chapter 4).

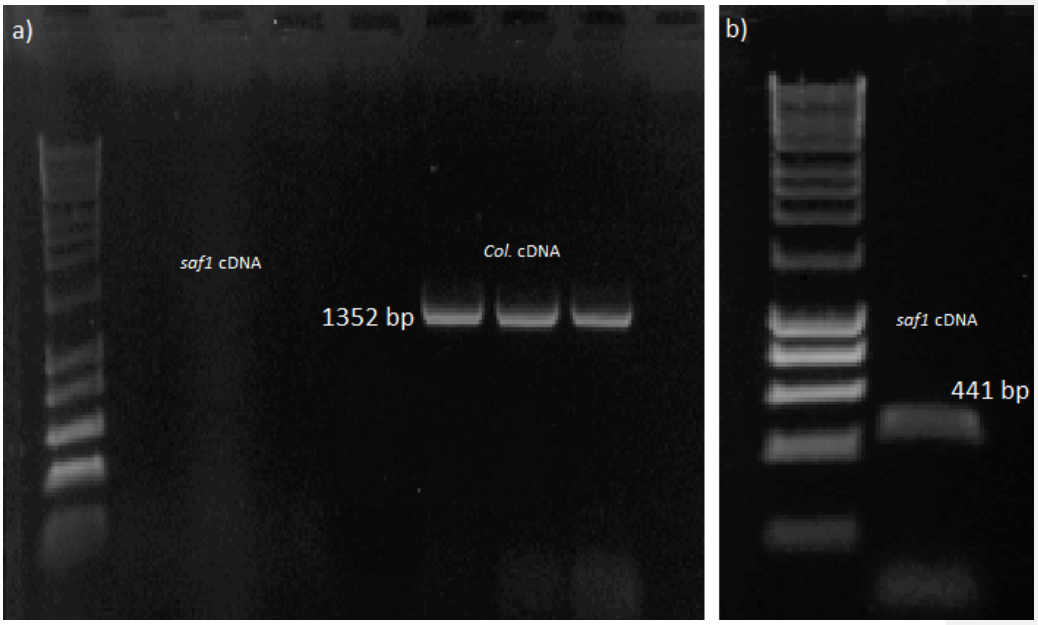

Figure 5.3 – a) PCR results for DNA amplification of *saf1* (SAIL_425_B06) cDNA and Col. cDNA using a whole gene primers (saf1_SAIL_LP and SAF1_qR-kim (Figure 5.2)) showing amplification of the gene in Col. DNA but not in the SAIL_425_B06). b) PCR results for DNA amplification of *saf1* (SAIL_425_B06) with qRT-PCR primers (SAF1_qF_kim with SAF1_qR-kim (Figure 5.2)). This confirms that the qRT-PCR primers amplify SAIL_425_B06 DNA and that the whole gene (a) was not transcribed in *saf1* mutants but was in wild type. Ladders in both gels are Hyper Ladder I.

To investigate how much of the gene was expressed in the *saf1* SAIL line amplification was carried out with a number different primers spanning across the gene (see Figure 5.2).

Using a reverse primer (SAF1_qR_kim) located at 1352 – 1373 bp with different forward primers (SAF1_SAIL_LP, saf1_exp_F2, saf1_exp_F and SAF1_qF_kim) were used to amplify SAF1 in the the wild type and SAIL_425_B06 insertion line. The wild type transcript was approximately 1352 bp, which equated to the

predicted length of the *SAF1* gene, however a partial transcript was observed from the *saf1* SAIL_425_B06 plants. Amplification of *SAF1* transcript was seen with all primer pairs in the wild type cDNA, whilst amplification was only seen in the *saf1* putative mutants using the SAF1_qF_kim / SAF1_qR_kim primer combinations (Figure 5.4), indicating that a partial transcript may be generated which starts somewhere between the SAF1_qF_kim and the saf1_exp_F primer to the end of the gene (441 - 776 bp (Figure 5.2)), suggesting that expression appears to be occurring downstream of the SAIL tDNA insert.

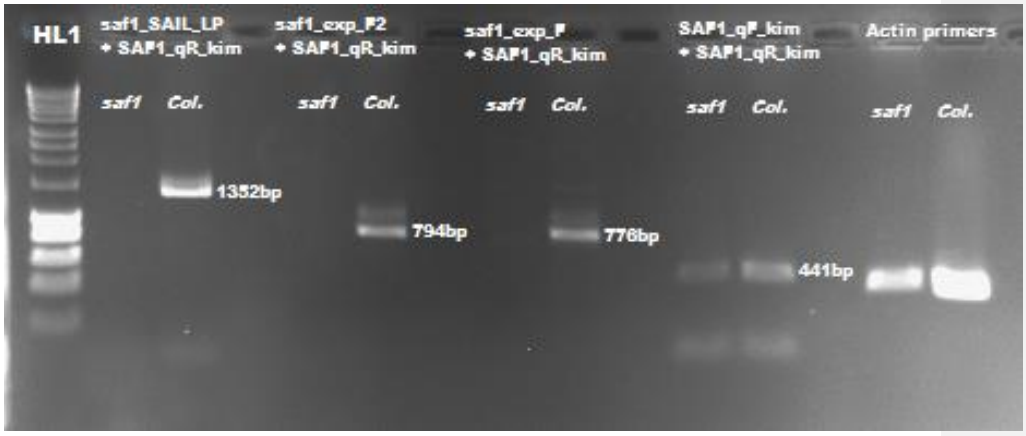

Figure 5.4 – PCR amplification of wild type cDNA and SAIL_425_B06 cDNA using primer pairs spanning different lengths of the *SAF1* gene. Wild type cDNA was successfully amplified throughout, whilst amplification of SAIL_425_B06 cDNA only starts for the last primer pair. This gives a start point for partial transcription of *SAF1* as somewhere between the location of saf1_exp_F and SAF1_qF_kim (Figure 5.2).

The transcribed sequence from the genomic *SAF1* DNA (obtained from TAIR) was translated using web.expasy.org (Figure 5.5) to identify the amino acid sequence of the SAF1 protein. Using http://smart.embl-heidelberg.de notable features (particularly the F-box domain and the FBD) of the expressed protein (exons from Figure 5.5) were identified (Figure 5.6). The partial transcript detected in the SAIL_425_B06 plants spans between the black markers in Figure 5.5, which means that the *saf1* knockout plants does not contain the F-

Box domain, but may contain the functional FBD domain. The FBD highlighted is a found in F-Box proteins but its exact feature is unknown (Doerks *et al*. 2002). The green box is the F-Box domain which is believed to be the important factor in ubiquitination (Bai *et al*. 1996). It seems unlikely this partial transcription leads to a functional F-box protein and so there may be redundant genes leading to a lack of phenotype associated with the *saf1* mutant (section 4.2.2). However, to determine whether the SAIL insertion line was due to the observed expression an alternative knock out line was generated using the CRISPR-Cas9 system.

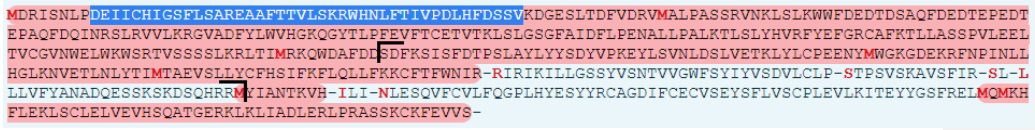

Figure 5.5 – Amino acid sequence of SAF1 protein. Amino acids highlighted in pink are the amino acids coded by the exons of the gene, whilst non-highlighted sequence is occurring in the introns and therefore should not be translated during actual expression. The blue highlighted amino acid sequence is the F-box domain. The section between the two black dividing lines contains the start point for translation for the *SAF1* partial translation that is detected in the relative gene expression in *saf1* plants. This translation was generated by web.expasy.org.

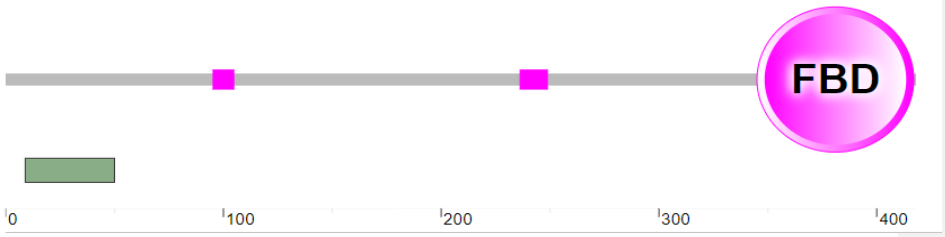

Figure 5.6 - Graphic representation of SAF1 protein from cDNA translation highlighting any notable features generated by *http://smart.embl-heidelberg.de*. The F-box domain is highlighted in Figure 5.6 as the green box, and as the blue highlighted amino acids in Figure 5.5.

5.2.2 Development of CRISPR saf1 knockout line

The *SAF1* insertional SAIL mutant line maintained partial expression of the *SAF1* gene, a *saf1* knockout line was therefore attempted to be generated using the CRISPR-Cas9 system.

Two sgRNA sequence for *SAF1* were transformed into an entry vector pCBC-DT1T2 (Section 2.6.1) which contained a guiding RNA scaffold (Figure 5.7). The relevant section containing the guiding scaffold and the *SAF1* specific sgRNA inserts was cloned into the Cas9 destination vector pBEE401e (Figure 5.8) through a Goldengate reaction (Section 2.6.2).

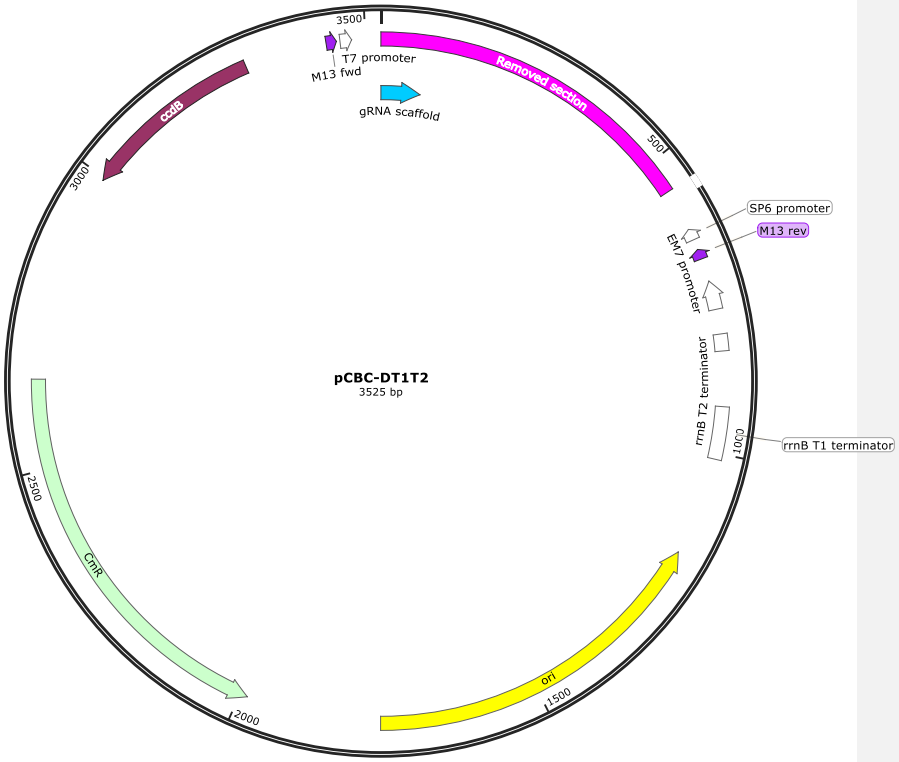

Figure 5.7 – A map of the entry vector that was used for the development of a knock out *saf1* plant line using a CRISPR/CAS9 system. The section of DNA that was removed from this vector and inserted into the destination vector pHEE-401e (Figure 5.8) is highlighted in the pink box labelled "removed section".

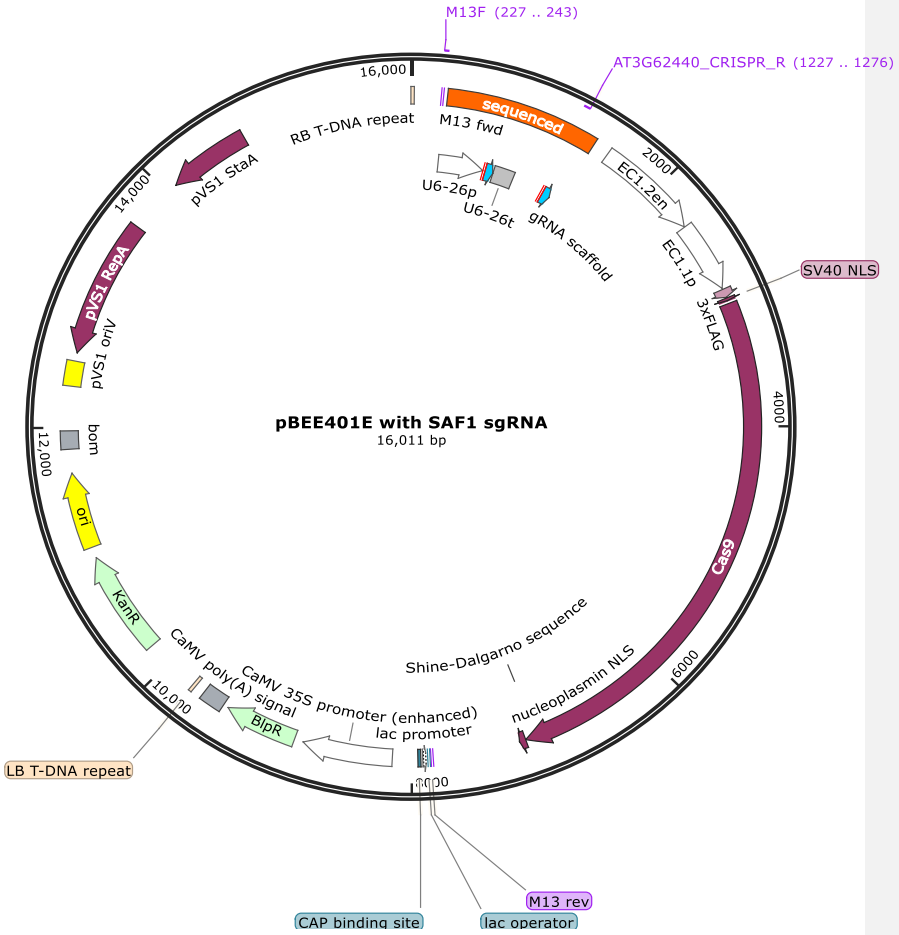

Figure 5.8 – A map of the Cas9 destination vector used for CRISPR knock out transformation. The orange section of the map has been successfully sequenced by Source Bioscience. The gRNA scaffolds which target the *SAF1* DNA are located are highlighted on this map and are within the sequenced section.

The pBEE401e vector containing the *SAF1* sgRNA inserts (the red section in Figure 5.8) was transformed into the *Agrobacteria* species gv3101 (Section 2.4.7). Colonies were tested for the correct insert of the *SAF1* sgRNA via a colony PCR (Section 2.4.4) with the primers AT3G62440_CRISPR_R and M13F (highlighted in Figure 5.8). PCR results suggest that the appropriate vector with the sgRNA inserts has been transformed successfully into gv3101 for individual

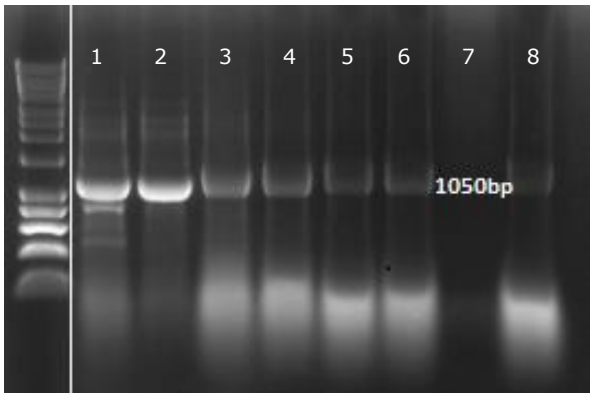

Figure 5.9 – PCR results for transformed agrobacteria colony PCR for 8 individual colonies. Primers used were AT3G62440_CRISPR_R and M13F to test for a *SAF1* sgRNA insert in the colonies.

colonies 1 – 6 and colony 8 (Figure 5.9). Colonies 2 and 3 (Section 2.1.3) were used to transform Landsberg *erecta* wild type plants to develop a CRISPR-Cas9 *saf1* mutant.

Putitatively transformed plants were initially grown on 1/2 MS nutrient plates without selection prior to transfer to soil and BASTA selection. Unfortunately germination of the putatively transformed seeds failed and did not survive to the selective stage. Further work to establish these mutant lines is therefore needed.

5.2.3 Possible Redundant Genes

5.2.3.1 Identifying Similar Genes

Using online gene sequence analysis software (https://salad.dna.affrc.go.jp/CGViewer/en/) a phylogenetic tree was generated for *SAF1* (Figure 5.10). The two most closely related genes, AT3G58920 and AT3G58960, were investigated further. Initially the expression localisation was investigated using the Kleptikova Atlas (Klepikova *et al*. 2016). Localisation of expression of AT3G58960 (Figure 5.11c) was very similar to *SAF1* (Figure 5.11a). AT3G58920 (Figure 5.11b) has expression different to *SAF1* with is being even more specific in its localisation, but it is localised to anther development during the same developmental stage (12-14) when expression of *SAF1* and AT3G5860 increases.

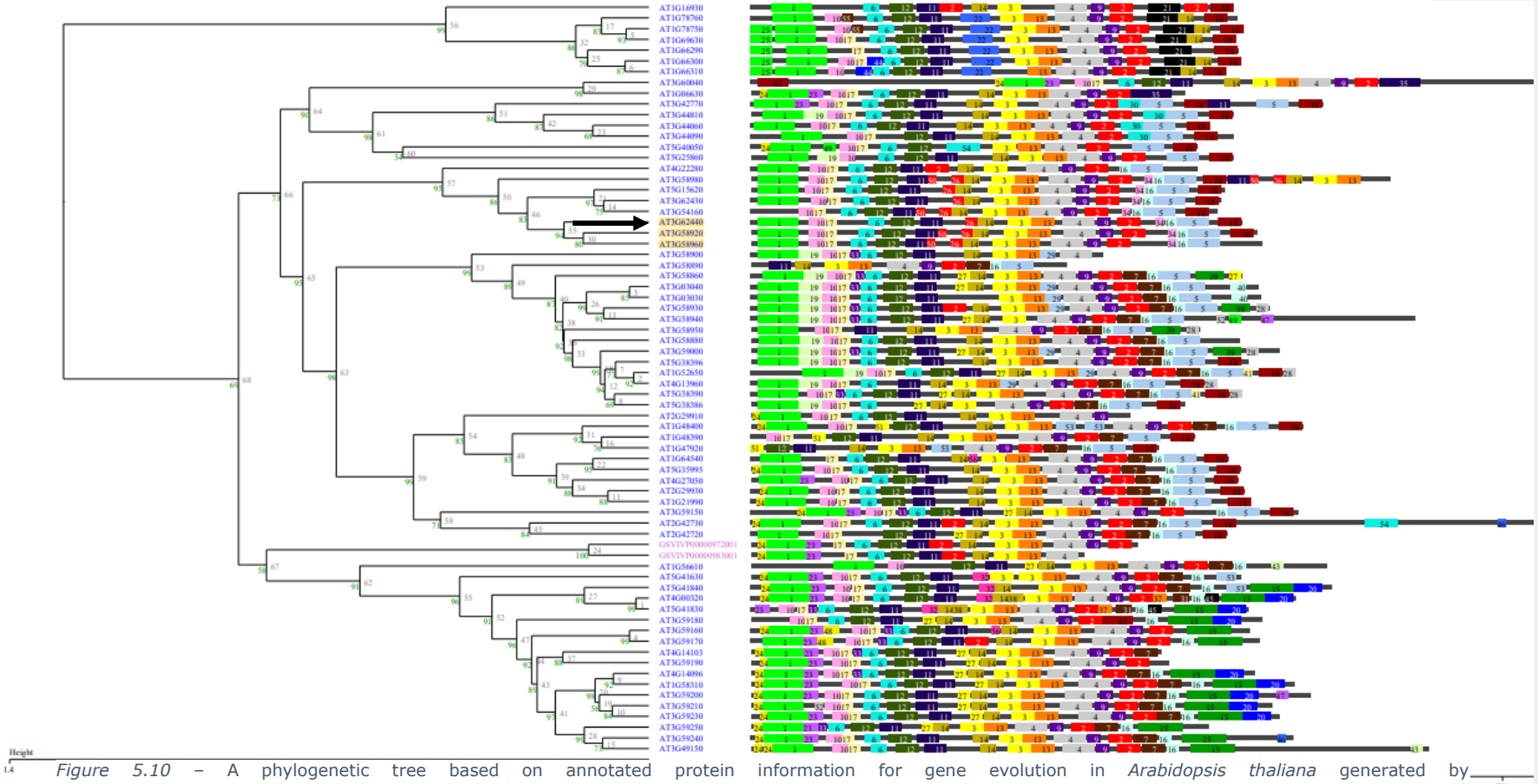

Figure 5.10 – A phylogenetic tree based on annotated protein information for gene evolution in *Arabidopsis thaliana* generated by https://salad.dna.affrc.go.jp/CGViewer/en/. *SAF1* (ATG362440) is highlighted along with the two most closely linked genes which were investigated further. The arrow highlights *SAF1.* Coloured boxes are different motifs with the green number 1 motif being an F box domain.

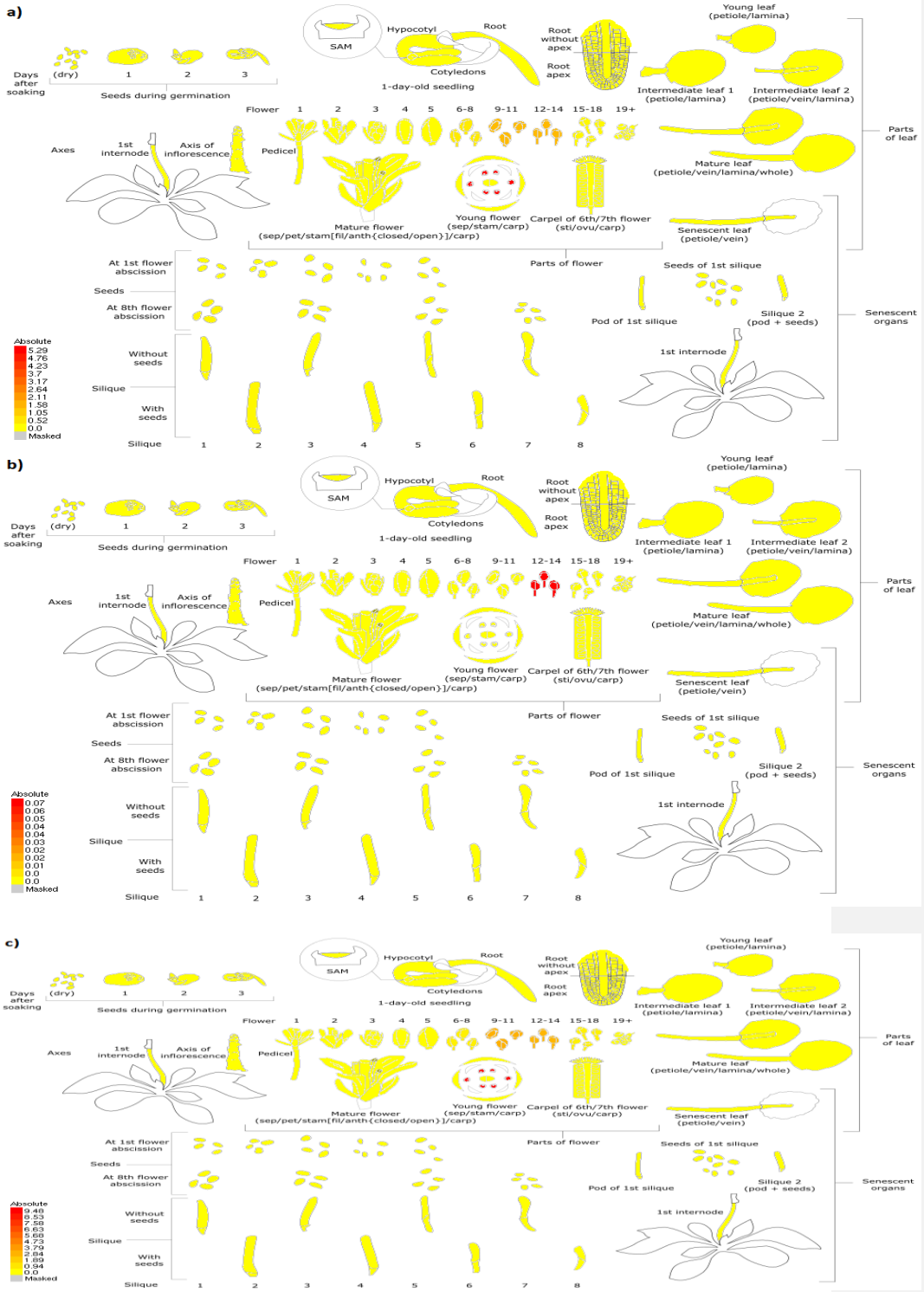

Figure 5.11 – Expression localisation of a) *SAF1,* b) AT3G58920 and c) AT3G58960 shown in Kleptikova Atlas (Klepikova et al. 2016).

The expression timing and tissue specificity was confirmed by qRT-PCR analysis. *Arabidopsis* wild type buds were collected and separated by size into 4 different stages. Buds were staged by size following protocols developed previously in the laboratory (Figure 5.12a) (Mo 2017). The expression of *SAF1*, AT3G58960 and AT3G58920 were investigated across the stages. Expression of all of these genes was relatively high in early stage buds before declining in older stages (Figure 5.12). Given that *SAF1* is removed from developing anthers to allow for the subsequent development of secondary thickening, possibly via the increase of NST1/NST2, and a similar expression of AT3G58960 and AT3G58920 is observed, it may suggest that they have a redundant role in anther dehiscence and endothecium thickening.

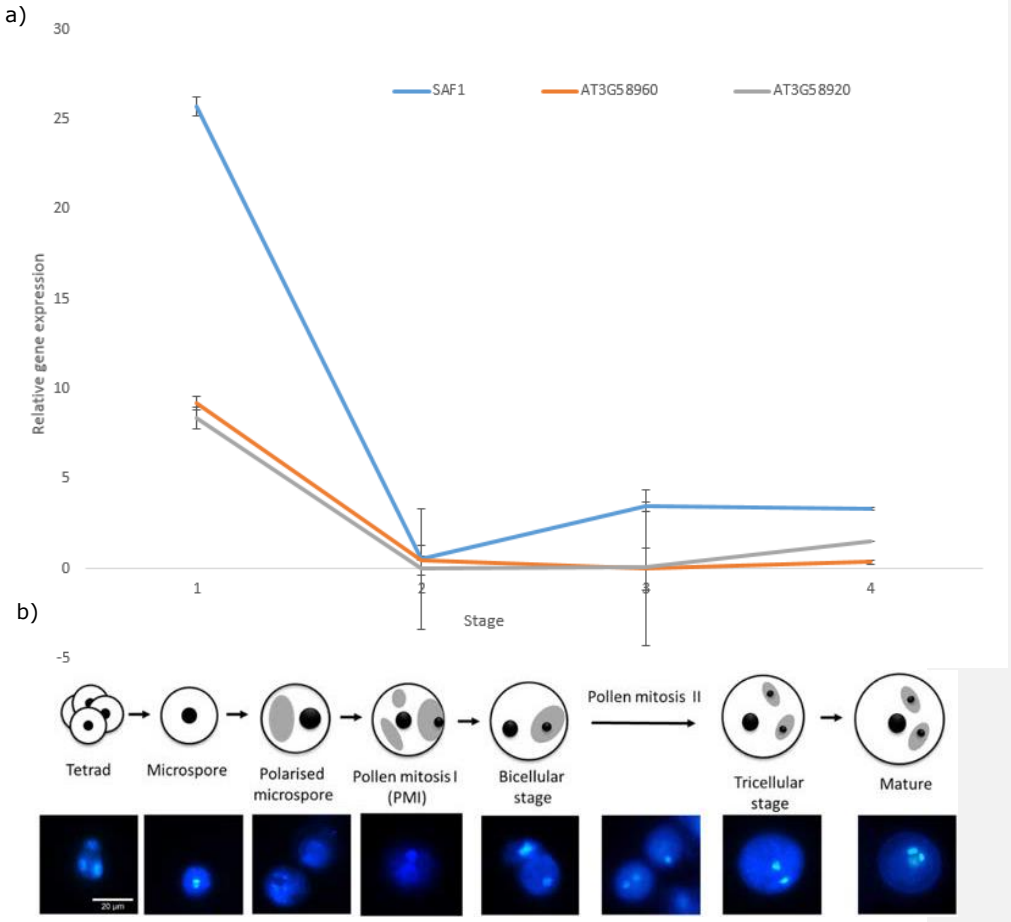

Figure 5.12 – a) Relative expression analysis in staged buds of *Arabidopsis thaliana* for *SAF1,* AT3G58960 and AT3G58920. There were 3 technical repeats of 3 individual plants for each set of staged buds. Error bars are standard error. Buds were staged by size based upon Mo (2017) (b) Stage 1 corresponding to polarised microspores, Stage 2 corresponding to the Bicelluar stage, Stage 3 corresponding to the Tricellular Stage and stage 4 were mature buds. An ANOVA was carried out with there being no significant ($p>0.05$) difference between plants AT3G58960 and AT3G58920 at any given stage, and no significant difference between *SAF1* and the other plants at stages 2-4, but there was a significant ($p<0.05$) difference between stage 1 and stages 2-4 in each plant line.

5.2.3.2 Phenotyping orthologue knock out lines

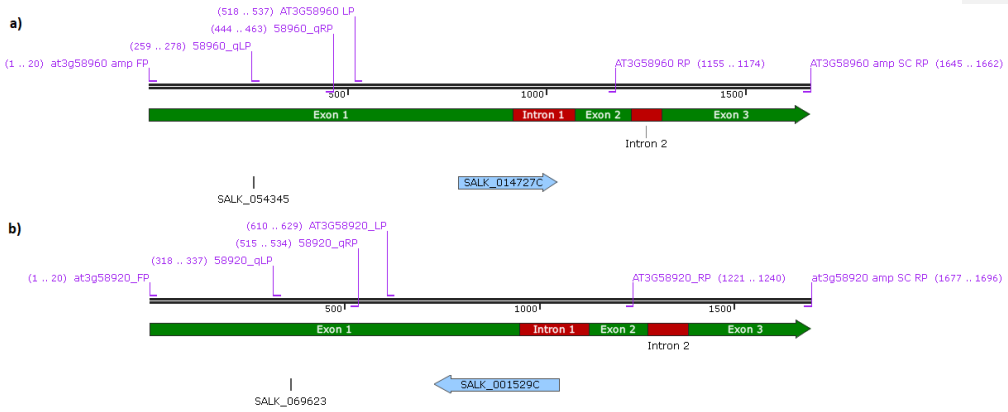

Figure 5.13 – Maps of a) *ATG358960* and b) *AT3G58920* highlighting the location of primers and the T-DNA SALK inserts.

Knock out lines for the possible redundant genes were ordered (SALK_054345C, SALK_014727C (AT3G58960), and SALK_069623C (AT3G58920)). These lines were genotyped (Section 2.3.3), with SALK_054345 plants being homozygous for the knock out insert (Figure 5.14a) whilst SALK_014727 did not appear to have the insert. All of the SALK_069623 (Figure 5.14b) plants were heterozygous for the tDNA insert. Plants 1-3 were used from SALK_054345C (homozygous) and SALK_069623 (heterozygous) going forwards for phenotyping and expression analysis. Lines were genotyped using the primers LBb1.3 with AT3g58960_RP and AT3G5860_LP or at3g58960_amp_FP (for SALK_014727 or SALK_054345 respectively). Whilst it is not ideal to work with the heterozygous plants, this was the only option available at the time, with the hope being that plants with one copy of the tDNA insert may have reduced expression and could give an indication of the effects of this. Ideally this would be repeated with homozygous plants.

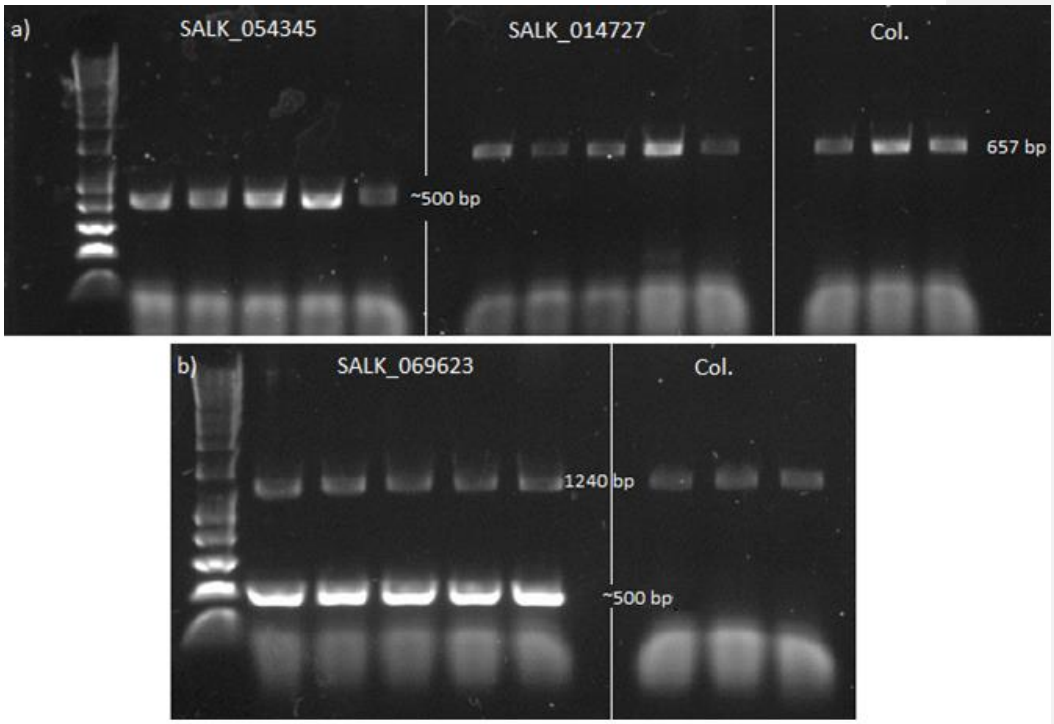

Figure 5.14 – PCR results for genotyping of 5 individual a) AT3G58960 SALK, b) AT3G58920 SALK and wild type plants. Primers were used in triplicate with two primers spanning the insert within the gene and primer LBb1.3 which is located within SALK inserts. This means AT3g58960_RP and AT3G5860_LP or at3g58960_amp_FP and 58960_qRP (for SALK_014727 or SALK_054345 respectively), and AT3G58920_FP with AT3G58920_RP for SALK_069623 (Figure 5.13). Ladder is Hyperladder I.

The siliques of plants 1-5 of *at3g58960* (SALK_05345) and 1-3 *at3g58920* (SALK_069623) were measured and compared to wild type plants. Ideally homozygous plants would have been investigated, however there were only heterozygous plants available and time constraints required these plants to be the investigated ones. There was no significant difference between the SALK lines and wild type silique lengths (Figure 5.15) suggesting that these SALK lines do not have a reduced fertility compared to wild type. Plants were then examined under a dissecting microscope. Both SALK_05345 (AT3G58960) and SALK_06923 (AT3G58920) knock out lines did not show any notable differences

(Figure 5.17 and Figure 5.18 respectively) to wild type (Figure 5.16) with pollen released from the developing anthers without any external agitation. Pollen was viable with Alexander staining in both SALK lines as seen in wild type plants.

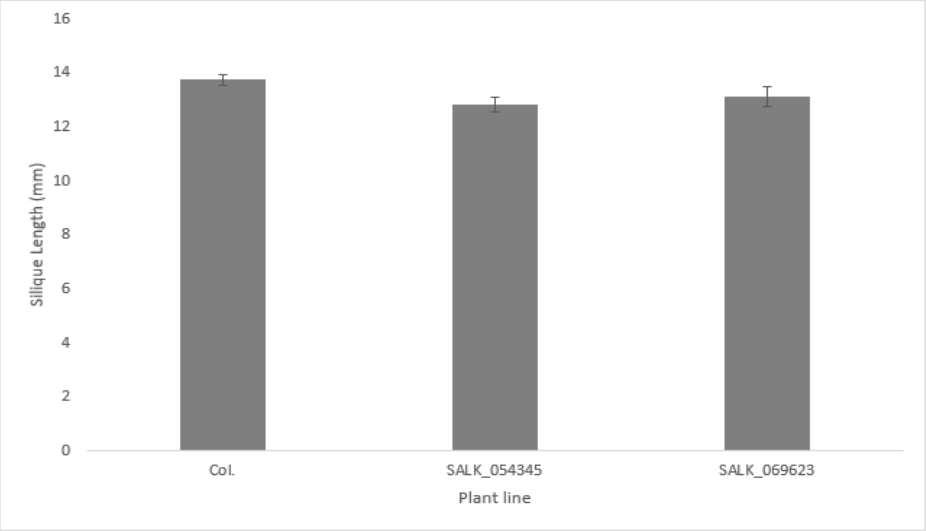

Figure 5.15 – Average silique lengths of wild type Columbia plants compared to SALK lines of potential *SAF1* redundant genes AT3G58960 (SALK_054345) and AT3G58920 (SALK_069623). There were 5 siliques measured on 5 individual plants for wild type and SALK_054345 and 3 individual plants for SALK_069623. Error bars are standard error.

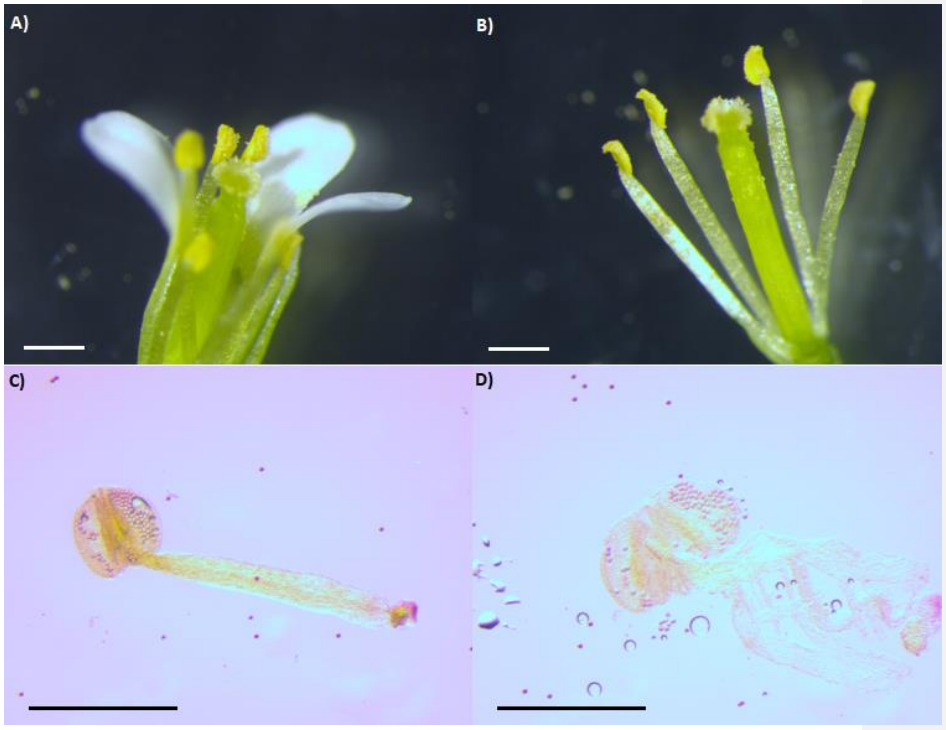

Figure 5.16 – Col. flowers dissected to highlight anther development. A) front petals removed to allow visualisation of the stamen. B) the same flower with the rest of the petals removed. Alexander staining was carried out to test for pollen viability with C) showing intact anther whilst D) shows the anther after it was lightly pressed to release the pollen. Scale bars are 1mm. This is the same image as Figure 4.7, but is reshown here for easier comparison with Figure 5.17 – Figure 5.18.

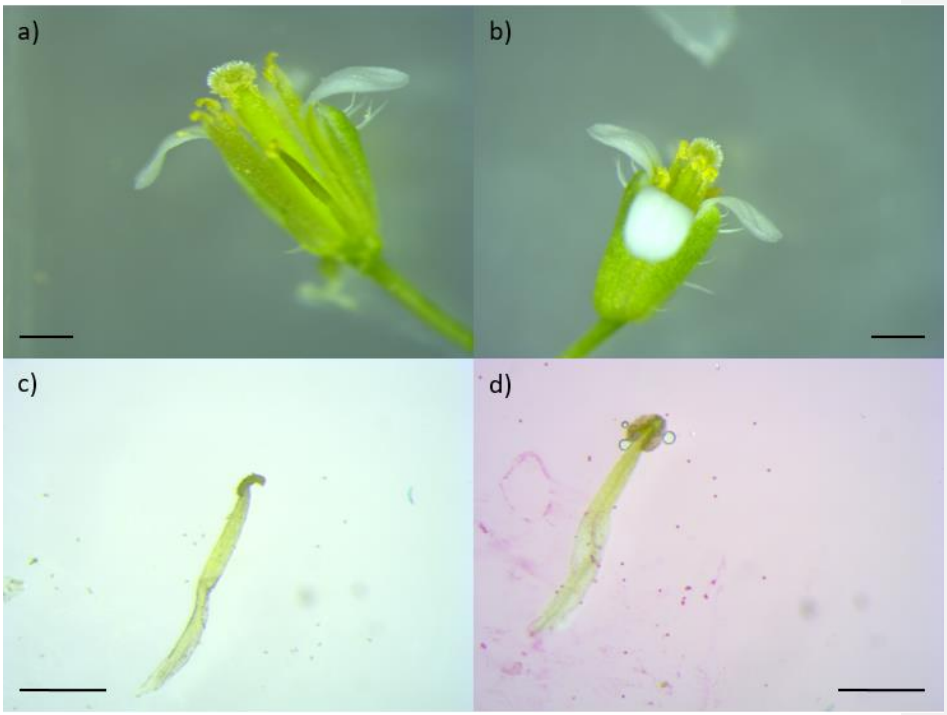

Figure 5.17 – Images of homozygous SALK_054345 (AT4G58960) plants (a & b), c) an anther removed from the developing flower and d) after Alexander staining to identify viable pollen. Scale bars are 1mm.

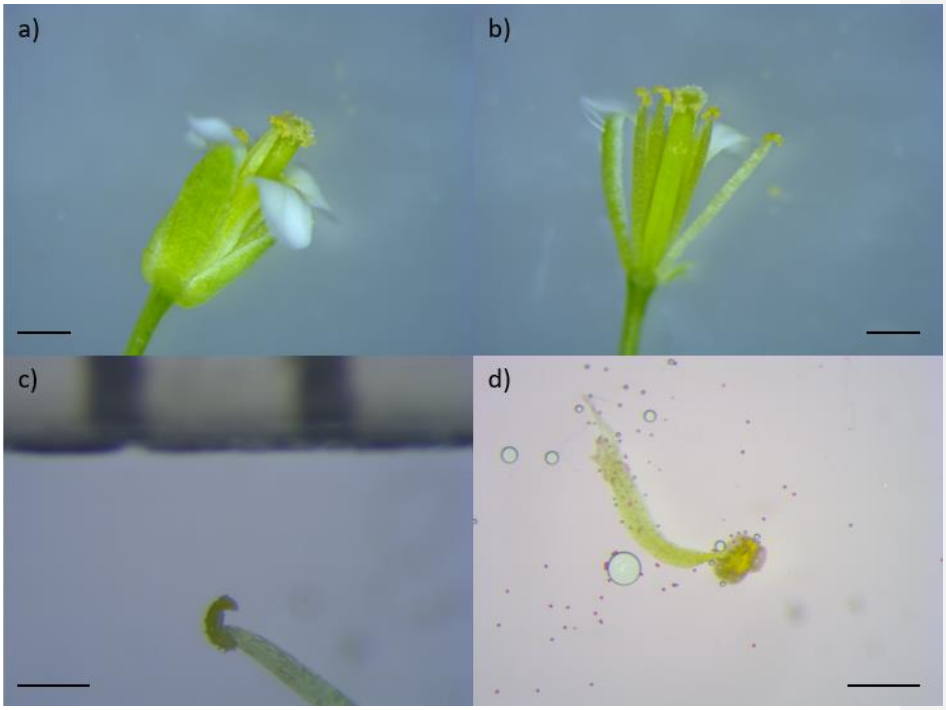

Figure 5.18 – Images of hetrozygous SALK_069623 (AT3G58920) plants (a & b), c) an anther removed from the developing flower and d) after Alexander staining to identify viable pollen. Scale bars are 1mm.

5.2.3.3 Expression analysis of orthologue genes

To investigate how expression of these possible *SAF1* redundant genes affects expression within the network, RNA was extracted (Section 2.3.1) from SALK_054345 and SALK_069623 and complimentary DNA was synthesised (Section 2.3.2). Relative gene expression of *MYB26, NST1, NST2, SAF1, IRX1, AT3G58960* and *AT3G58920* was investigated in Col. wild type, SALK_05345 (AT3G58960) and SALK_069623 (AT3G58920) lines (Figure 5.19).

In both the SALK lines there was complete knock out (no significant difference ($p<0.05$) to no expression) of the putative orthologue gene of interest. When *AT3G58920* was knocked out, the *AT3G58960* also appeared knocked down, whilst knocking out *AT3G58960* did not have a significant impact on *AT3G58920* expression. Knocking out *AT3G58920* did not lead to a significant change in

MYB26 expression or the downstream genes, but there was a trend towards reduced expression of *NST1, NST2* and *IRX1*. Interestingly this was different to knocking out *AT3G58960,* whilst *ATG3G58920* expression was not significantly changed (SALK_05345). In this scenario, when *AT3G58960* expression was reduced, *MYB26* expression increased (3.74x compared to the wild type expression) which in turn led to a significant increase in the expression of *IRX1* (3.82x wild type expression). *NST1* expression was lower (0.71x) in *AT3G58960* knock out plants and *NST2* expression was higher (1.38x), but not to a statistically significant level.

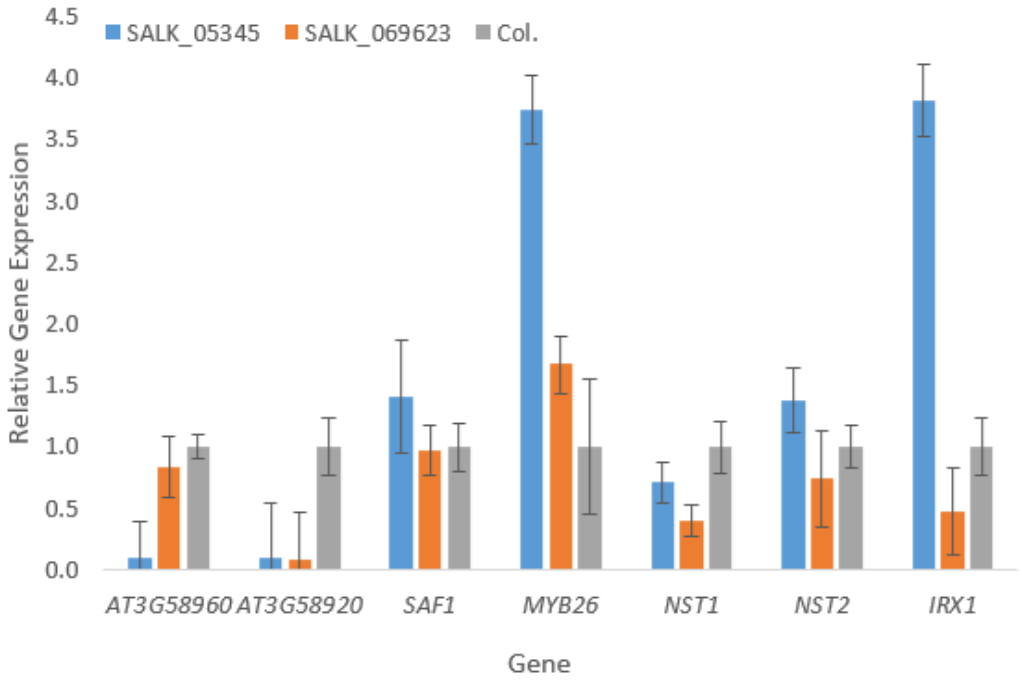

Figure 5.19 – Relative gene expression of a number of genes in *AT3G58960* knock out plants (SALK_05345) and *AT3G58920* (SALK_069623) compared to wild type expression in wild type Col. plants.

5.2.4 Spatial and Temporal Localisation of other Genes

Evaluation of the localisation of *TGA9* (Figure 5.20), *TGA10* (Figure 5.21) and *PKSP* (Figure 5.22) and *PCRK1* can be investigated to see if it aligns with *SAF1,*

MYB26 and *NST1/NST2. TGA9, TGA10* and *PKSP* all have high expression in the developing anthers, particularly at the stages when anther endothecium formation is occurring. *TGA9* has the highest level of expression, whilst *TGA10* has very similar expression but at a lower level in the developing anthers. *PSKP* expression is high throughout developing plants. There is expression of *PCRK1* (Figure 5.23) in developing anthers but it is more highly expressed in leaf and stem development.

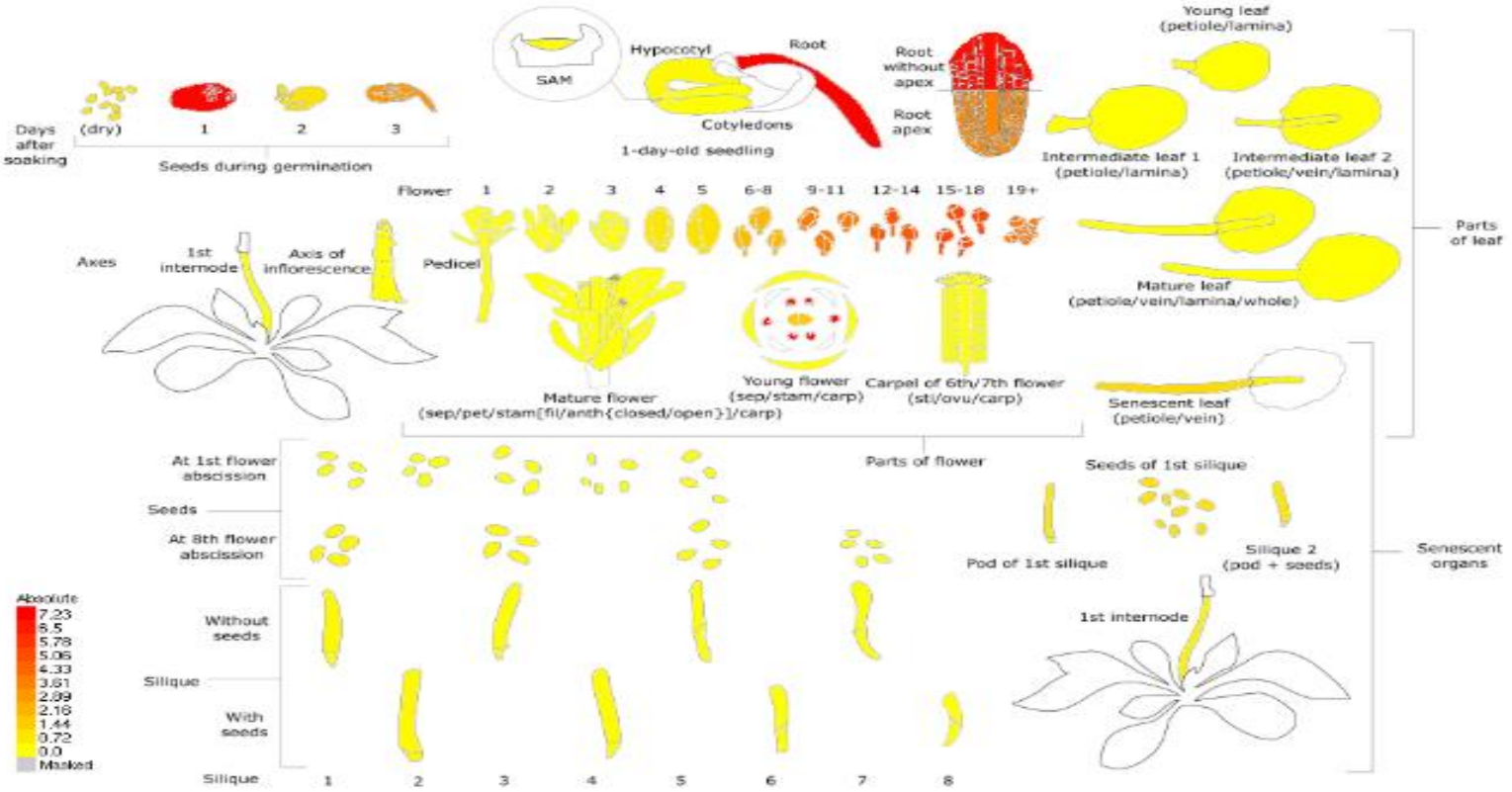

Figure 5.20 – Expression localisation of *TGA9* as shown in Kleptikova Atlas (Klepikova *et al*. 2016).

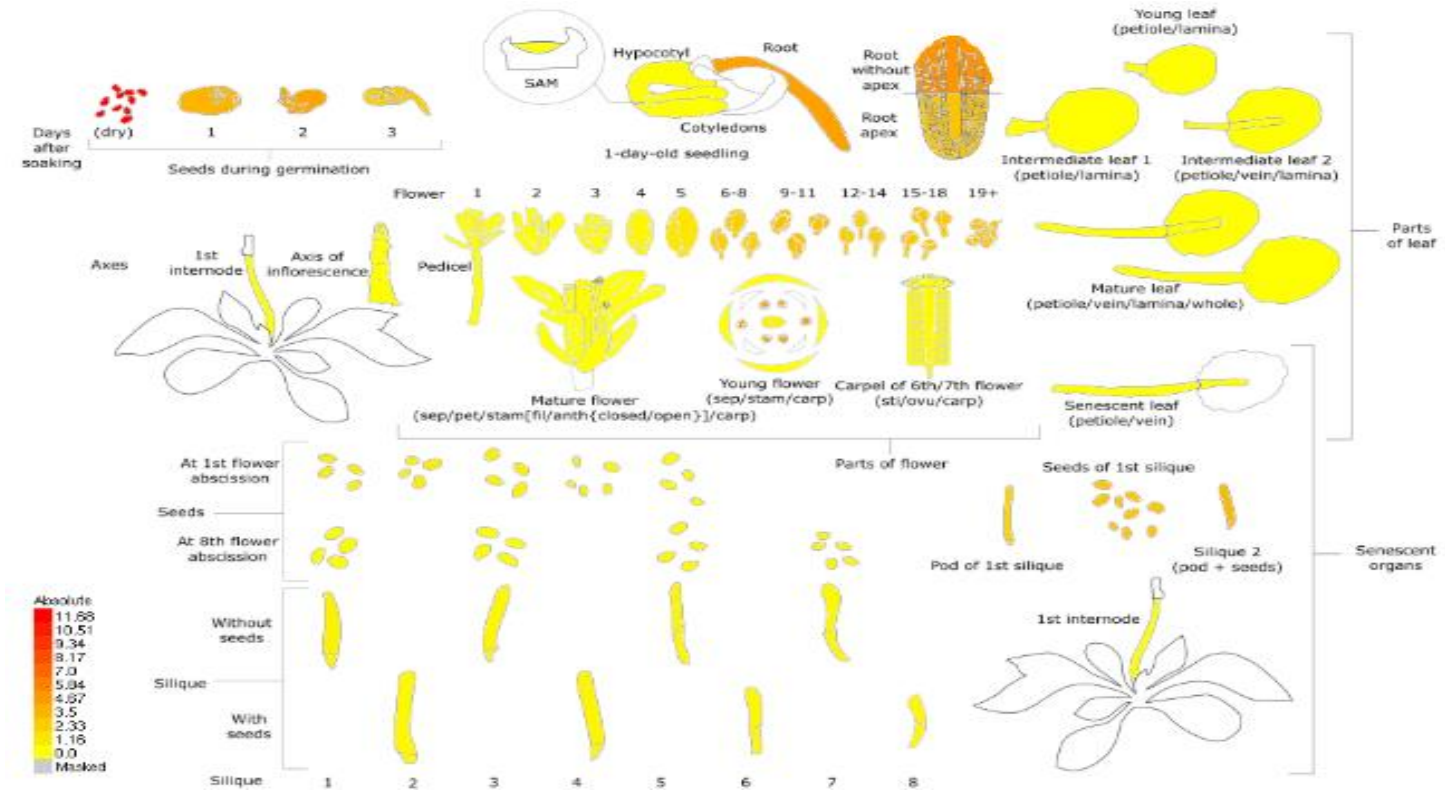

Figure 5.21 – Expression localisation of *TGA10* as shown in Kleptikova Atlas (Klepikova et al. 2016).

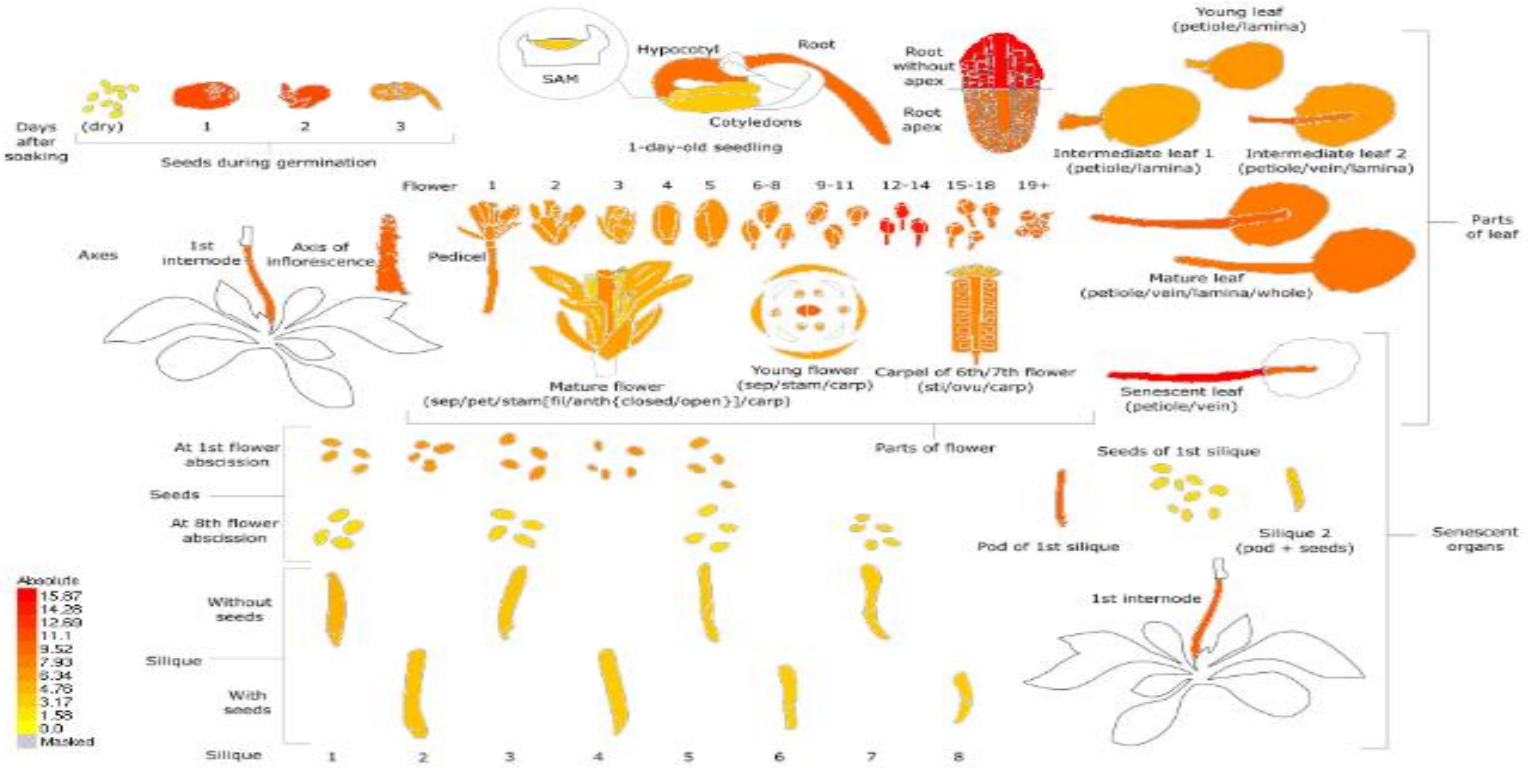

Figure 5.22 – Expression localisation of *PKSP* as shown in Kleptikova Atlas (Klepikova et al. 2016).

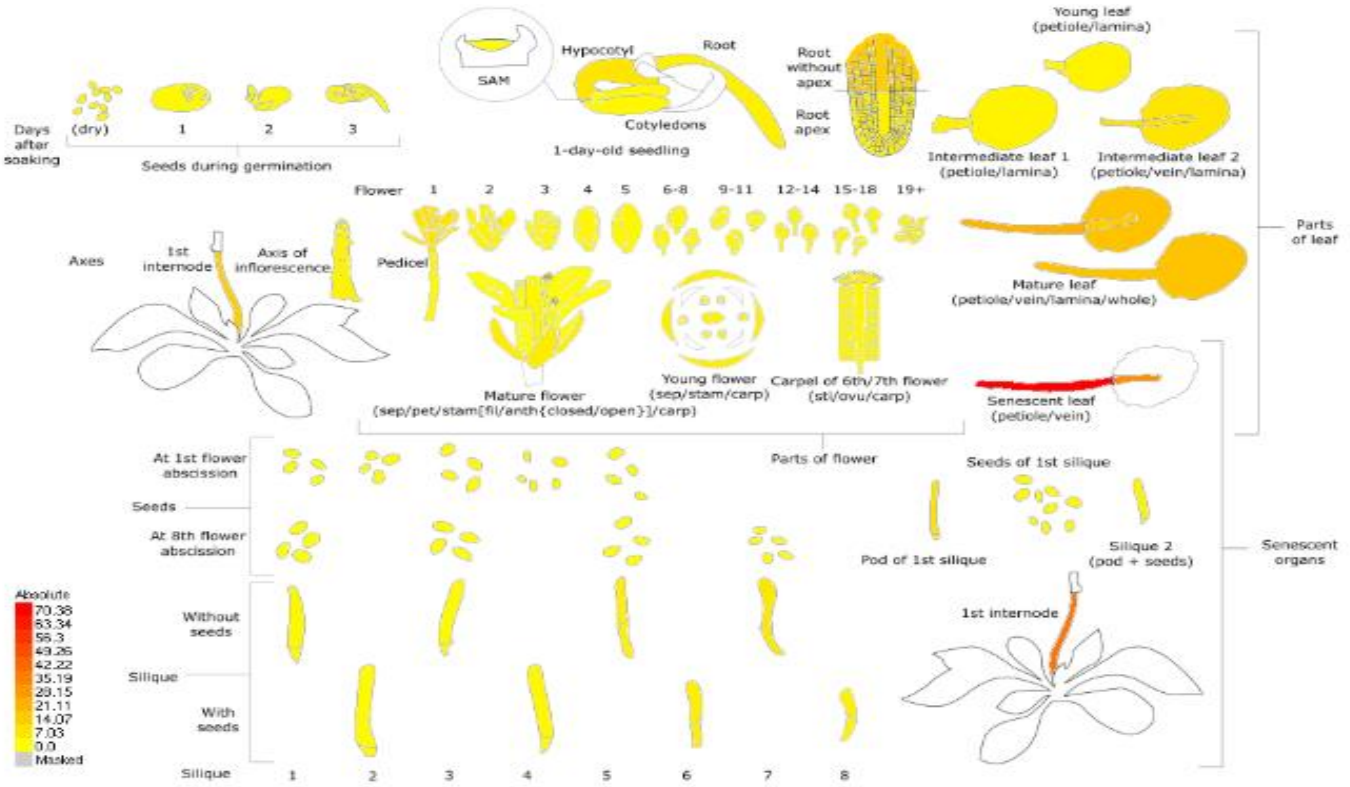

Figure 5.23 – Expression localisation of *PCRK1* shown in Kleptikova Atlas (Klepikova et al. 2016).

5.2.5 Effects of *SAF1* Overexpression on Potential *MYB26* Target Genes and Analysis of Expression of Known Network Genes in Mutants Lines for these Genes

The expression of *TGA9, TGA10, PKSP* and *PCRK1* in 35S::*SAF1* plants was examined. Figure 5.24 indicates that *SAF1* was significantly ($p<0.01$) overexpressed (15x compared to wild type) in these plants. The overexpression of *SAF1* led to a significant ($p<0.01$) downregulation of expression of *PKSP* and *TGA9.* there was no statistical significant difference between wild type and 35S::*SAF1* expression of *TGA10* and *PCRK1* using A Student T-test (Figure 6.2b).

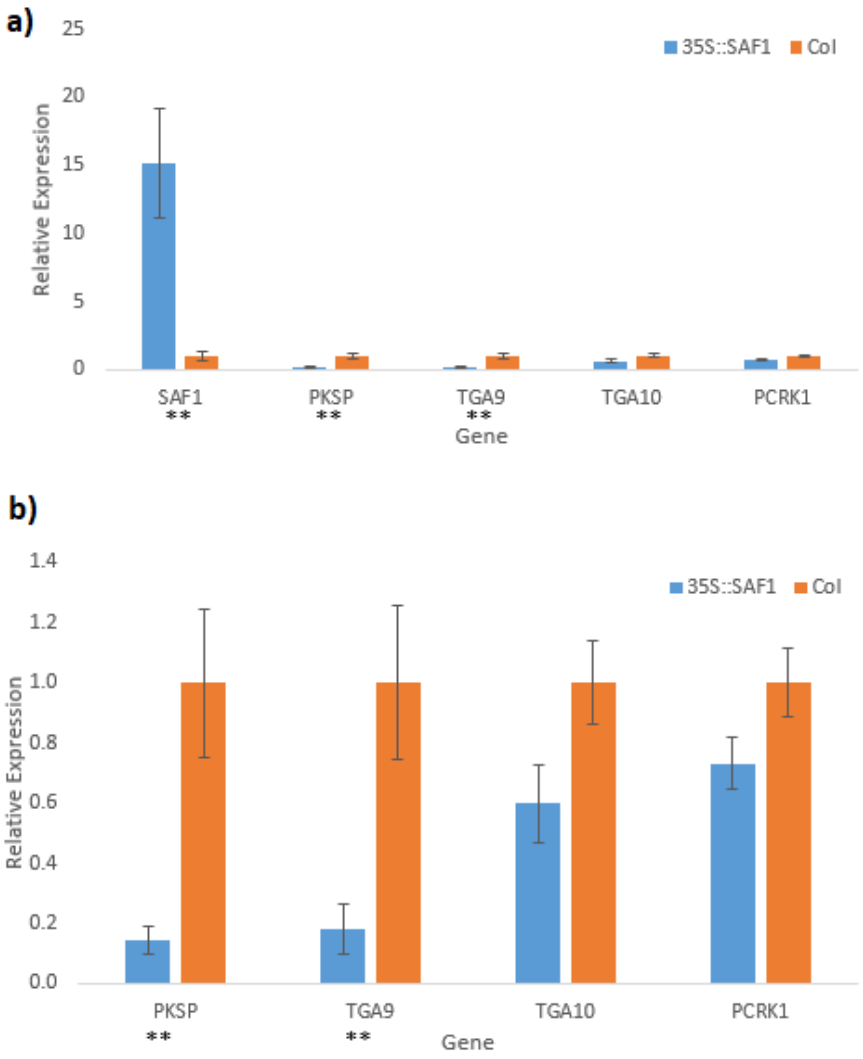

Figure 5.24 – The relative expression of some of possible *MYB26* target genes in 35S::*SAF1* overexpression line compared to wild type plants. a) contains all of the values, whilst b) removes the *SAF1* expression for easier visualisation of expression of other genes. Error bars are standard error. Statistical analysis was carried out using a Student T-test. There were 3 technical repeats of 2 biological repeats of 35S::*SAF1* (due to the number of healthy plants available to collect RNA from) and 3 technical repeats of 3 biological repeats of Col. wild type plants.

Quantitative RT-PCR (Section 2.3.5) was carried out on *tga9/tga10* line (SALK_057609 and SALK_124227) and a *nst1/nst2* line. The expression of potential *MYB26* target genes or interacting proteins was examined in these mutants (Figure 5.25). As previously seen (Section 4.3), the overexpression of *MYB26* effectively led to the knockout of *SAF1* (x0.04 expression), whilst knocking out *myb26* led to a large increase in *SAF1* (x3.9 compared to wild type) expression. Interestingly, both the overexpression and reduction of *MYB26* expression led to reduced expression of *NST1, NST2, TGA9, TGA10, PKSP* and *IRX1*. The double *nst1nst2* mutant had reduced expression of the genes downstream of *MYB26*.

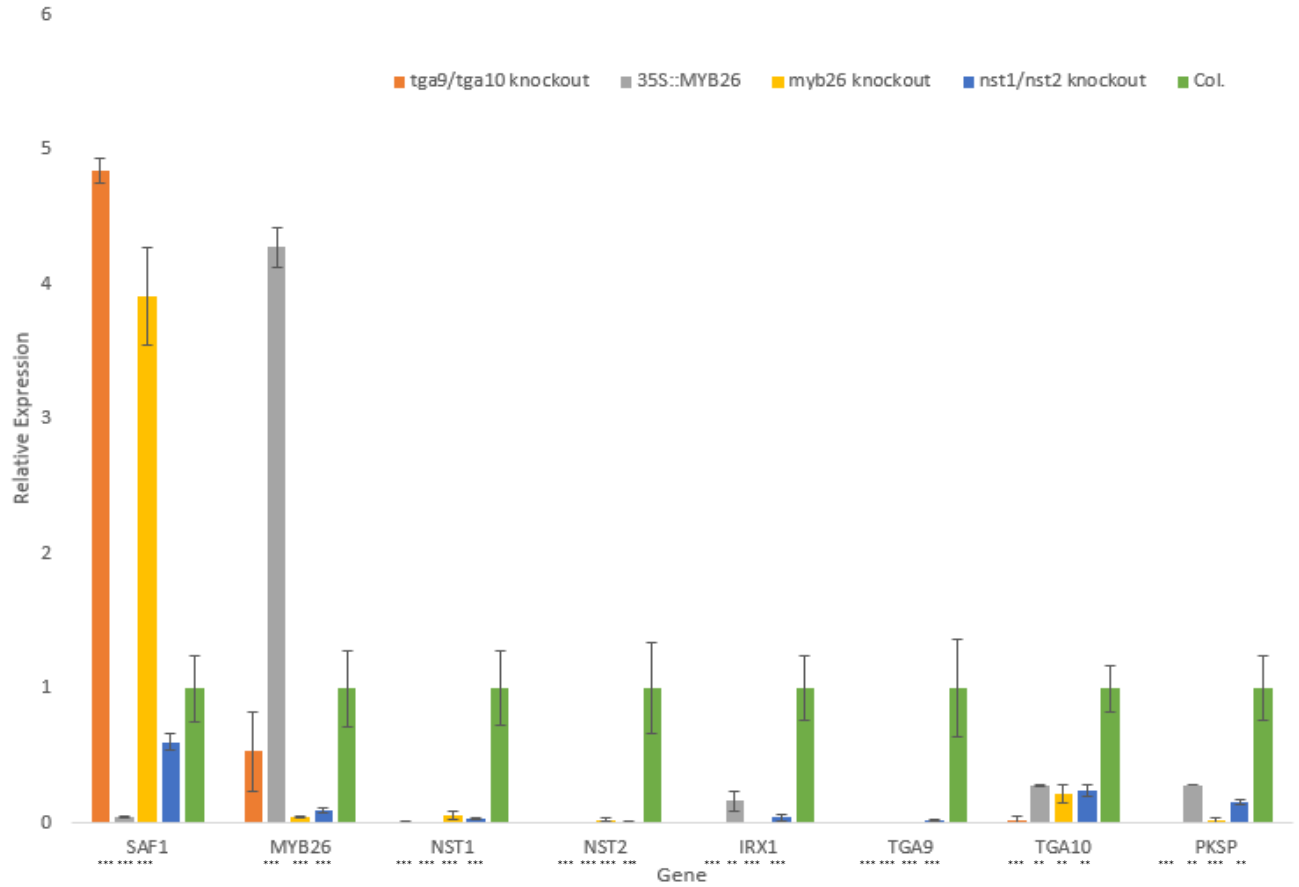

Figure 5.25 – The expression of a number of genes in different mutant lines, including a double knockout mutant for *tga9* (SALK_057609) and *tga10* (SALK_124227), an overexpression line for *MYB26,* a *myb26* (SALK_112372), and a double knockout for *nst1* (SALK_120377) *nst2* (SM_3_19668). Error bars are standard error. There were 3 technical repeats for 3 biological repeats for each line, except for 35S::*MYB26* and *myb26* knockout plants where there was only one biological repeat with 3 technical repeats. A Student T-test was done. To compare each expression level to wild type expression.

5.3 Discussion

5.3.1 *saf1* and Potential Redundant Genes

T-DNA inserts are regularly used to knock out genes in *Arabidopsis thaliana* (Feldmann 1991). Previously (Chapter 4) *SAF1* function was investigated using a SAIL insert T-DNA line to knock out *SAF1* expression. Unfortunately, expression of *SAF1* was detected in these SAIL_425_B06 plants. Expression of *SAF1* was analysed by testing potential amplification of *SAF1* across varying lengths of the gene (Figure 5.4) to help determine the extent of partial transcript produced from the *saf1* mutant. This partial expression of *SAF1* could lead to the translation to a partial protein being produced within the cell, however Figure 5.5 and Figure 5.6 highlight that this partial protein does not contain the F-box domain of *SAF1* associated with ubiquitination (Kim et al. 2012; Schumann et al. 2011). The lack of an F-box domain in partial SAF1 proteins would suggest that it would be non-functional, and therefore the absence of expression changes in downstream genes (which was observed in Chapter 4) is not due to the prescence of a functional SAF1 protein but is more likely to be due to redundant gene expression. However it must be considered that, the partial SAF1 protein contains an FBD motif (Figure 5.5), which is present in F-box proteins but has not, to date, had a function identified (Doerks et al. 2002) – the FBD motif may play a role in regulating anther endothecium thickening and so it is impossible to conclude that the partial protein is definitely non-functional.

Using phylogenetic analysis looking at the tree of gene evolution based on annotated protein information, the most closely related genes to *SAF1* were identified (Figure 5.10) and expression localisation of these genes was examined (Figure 5.11).

AT3G58960 and *AT3G58920* both encode F-box/RNI-like superfamily proteins (Cheng et al. 2017) similar to *SAF1*; this supports the potential redundancy of

these genes. It is therefore possible that these genes are all able to tag NST2 with ubiquitin leading to the removal of NST2 from the cells. When the phenotype of knock out lines for AT3G58960 and AT3G58920 (SALK_05345 and SALK_069623 respectively) were analysed, there was no notable difference between these plants (Figure 5.16 and Figure 5.17) and wild type plants (Figure 4.7), as found for *saf1* plants (Section 4.2.3). Expression of *MYB26* and downstream genes was investigated in the SALK_05345 and SALK_069623 plants and whilst both genes were knocked out in the SALK lines, there was no significant change of gene expression in *MYB26* or downstream genes in *AT3G58920* knock out, although *AT3G58960* expression was knocked out in this SALK_069623 line. Interestingly, in the *AT3G58960* knock out plants (SALK_05345) there was a significant change in gene expression despite *AT3G58960* also being knocked out in the SALK_069623 plants. This would suggest that when only *AT3G58960* is knocked out then there is a significant change in expression, whilst knocking out both *AT3G58960* and *AT3G58920* doesn't have any significant effect. Potentially knocking both genes out may lead to a compensatory action which does not occur in the *AT3G58960* knock out plants, although further investigation would be needed to develop a more detailed understanding of this.

In the SALK_05345 *MYB26* expression is increased. This could be the case if NST2 protein accumulated to a higher level than in the wild type as NST2 promotes *MYB26* expression which supports the idea that *AT3G58960* is involved in anther endothecium thickening in a similar role to *SAF1.* However, the much larger (~3.5x) increase in *MYB26* expression and *IRX* expression in the *AT3G58960* knock out plants compared to the non-significant increases in *NST1/NST2* expression could suggest that *AT3G58960* has a role in anther endothecium thickening via a separate pathway. Mo (2017) identified proteins that are potentially interacting with *MYB26* along with downstream genes, and it may be that some of these newly identified related genes and proteins are

involved in a separate pathway for anther endothecium thickening which potentially involves these two F-box/RNI-like superfamily proteins.

5.3.2 *MYB26* Downstream and Putative Interacting Proteins

The expression localisation of *TGA9* (Figure 5.20)*, TGA10* (Figure 5.21) and *PKSP* (Figure 5.22) is relatively high in developing anthers, with late stages having a particularly high expression level. Since anther endothecium secondary thickening occurs around stages 11 onwards (Goldberg *et al.* 1993; Smyth *et al*. 1990) it is possible that these genes may play a role in the biosynthesis of secondary cell walls in the endothecium of developing anthers (Murmu et al. 2010).

PCRK1 is a gene with a similar sequence to *PKSP* (Mo 2017; Sreekanta *et al.* 2015) and so may play a similar role. However, the expression of *PCRK1* (Figure 5.23) is relatively high in developing leaves, with only a low level of expression in developing anthers. It could be that, similar to *NST1* that it is more important in secondary cell wall biosynthesis in vascular tissues, whilst *NST2* is more important in secondary cell wall biosynthesis in developing flowers (Mitsuda *et al*. 2007; Mitsuda *et al*. 2005), *PCRK1* could have similar function to *PKSP* but in a wider/different range of tissues whilst *PKSP* is more anther specific.

The suggestion that *TGA9* and *PKSP* play a role in anther endothecium thickening is strengthened by the fact that *SAF1* overexpression leads to a statistically significant downregulation of these two genes (Figure 5.24). It is known that overexpressing *SAF1* leads to male sterile plants due to a failure of endothecium secondary thickening due to a reduction in secondary cell wall biosynthesis genes (Kim *et al.* 2012). Work presented in Chapter 4 suggests that overexpression of *SAF1* leads to a downregulation in the expression *NST1, NST2* and *MYB26* – this could also be due to reduction in NST2 accumulation which in turn causes a downturn in *MYB26* expression and all downstream genes. If *TGA9* and *PKSP* are positively regulated by *MYB26,* as suggested by Mo (2017), then the overexpression of *SAF1* reducing the accumulation of NST2

would lead to the reduced expression of *TGA9* and *PKSP,* as observed here. *TGA10* and *PCRK1* both have a non-significant reduced expression in the *35S::SAF1* line, which considering these genes are less specific to developing anthers, could be due to the fact they have a redundant similar function to their relevant orthologues in anthers, but are less tissue-specific and therefore are less important in developing anthers.

Mo (2017) phenotyped *tga9/tga10* knockout lines previously and noted that they were male sterile due a lack of pollen release. To investigate if this may be due to a breakdown in the *MYB26* controlled network for anther endothecium thickening, relative gene expression in this mutant was investigated. As with *myb26* plants (Section 4.3), knocking out both *tga9/tga10* led to an increase in the expression of *SAF1* (Figure 5.25). It could therefore be that *MYB26* downregulates *SAF1* expression through the activation of *TGA9/TGA10* (Figure 5.26a).However, in the *35S::MYB26* line, there is a decrease in the expression of *TGA9* and *TGA10.* Mo (2017) identified TGA9 as potentially interacting with MYB26 rather than necessarily a direct downstream activation, and possibly TGA9 (and potentially TGA10) and MYB26 interact to decrease the expression level of *SAF1* (Figure 5.26b).

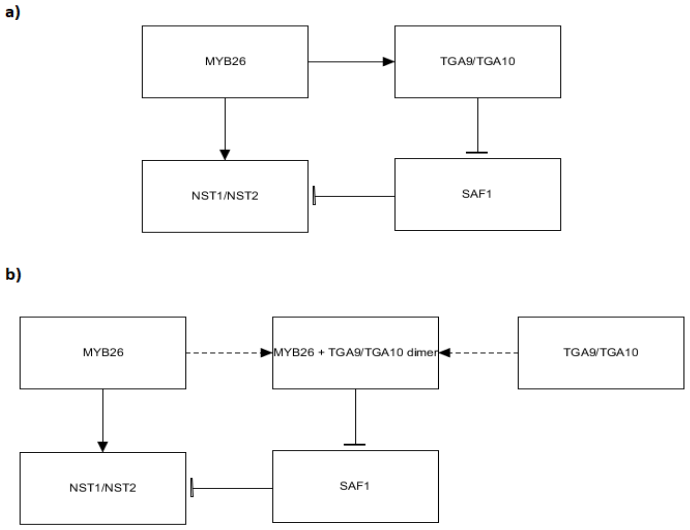

Figure 5.26 – Potential ways *TGA9* and *TGA10* may interact with *MYB26* to downregulate *SAF1* expression. a) suggests *MYB26* directly upregulates *TGA9/TGA10* which in turn downregulates *SAF1.* b) is an alternative model where TGA9 and/or TGA10 interact with MYB26 to downregulate *SAF1.* Full arrows are upregulation, dashed arrows are protein production and bar represents repression (either by inhibiting gene expression or by preventing protein accumulation.

However, it should be noted that these results (Figure 5.25) are unexpected because the *35S::MYB26* led to a downregulation of all downstream genes such as *NST1/NST2* and *IRX1* which does not correlate with what would occur if there is a simple activation of *NST1/NST2* from MYB26. Previous work has suggested that overexpression of *MYB26* leads to a subsequent increase in *NST1/NST2* expression (Yang *et al*. 2007) and therefore would be worth repeating this experiment with a new *MYB26* overexpression line.

Knocking out both *nst1* and *nst2* led to a downregulation of *MYB26,* as shown in previous work (Yang *et al*. 2017), and all subsequent downstream genes, which would be expected and fits the network model developed previously (Chapter 3).

With regards to *PKSP* expression there is downregulation in all the mutant lines examined above. This corresponds to the suggestion that *PKSP* is directly upregulated by *MYB26* in the *nst1/nst2, tga9/tga10* (due to an increased *SAF1* expression) and *myb26* lines. However, the *35S::MYB26* plants have results different to what would be expected if *PKSP* is upregulated by *MYB26* through the direct binding of MYB26 to *PKSP* (Mo 2017), which would suggest overexpression of *MYB26* should lead to an increased *PKSP* expression.

5.3.1 Future work

There are a number of experiments which would be interesting to carry out to investigate the knocking out of *saf1* along with the potential redundancy and genetic orthologues. A CRISPR vector for *saf1* knock out plants has been generated, but unfortunately transformed seeds did not germinate, further analysis of these lines is needed to confirm the function of the *SAF1* gene. Once CRISPR knock out plants are generated then expression analysis similar to Section 4.2.3 could be carried out to see if a lack of *saf1* expression led to similar relative gene expressions as in the T-DNA insertion mutants.

Future investigation of the potential *SAF1* orthologues is important. Ideally investigating relative expression in overexpression lines for *AT3G58960* and *AT3G58920* would be carried out to see along with phenotyping to compare to 35S::*SAF1* plants. If genes are working redundantly, it would be very interesting to generate a triple knock out (*saf1/at3g58960/at3g58920*) to carry out relative gene expression and phenotypic analysis on. Since these genes encode F-box/RNI-like superfamily proteins, as with *SAF1,* it is likely that any interactions occur on a protein – protein level. Investigating protein – protein interactions through methods such as Yeast-2-Hybrid, FRET or luciferase assays

to see if AT3G58960 or AT3G58920 interacts with MYB26, NST1, NST2 or any of the genes identified by Mo (2017) would be useful to gain a more precise understanding of *SAF1* and the role of potential redundant genes.

Due to the results from the *35S::MYB26* (Figure 5.25) being very different to the expected results when compared to network hypotheses and previous work (Mo 2017; Yang *et al.* 2017) and the results based on a small n number (due to only small number of biological replicates growing successfully) it would be important to repeat this work. In addition to this, to understand the interaction of these additional genes in the network, overexpression lines for *TGA9* should be investigated to see how this affects the expression of the genes described in Figure 5.25. A *pksp* line and an overexpression line for *PKSP* should be investigated to further understand *PKSP*'s potential role within the network.

Mo (2017) also identified another gene which appeared to be upregulated by *MYB26* called *Y2H620* (*TCP14*). It would be worth investigating this gene further including expression localisation, relative expression within the mutant lines discussed here, along with developing mutant and overexpression genes for *TCP14.* An additionally class II TCP family protein TCP24 has also been identified negatively regulating secondary thickening in anthers and roots and it would be interesting to also investigate how this protein may fit into the network.

Chapter 6: General Discussion

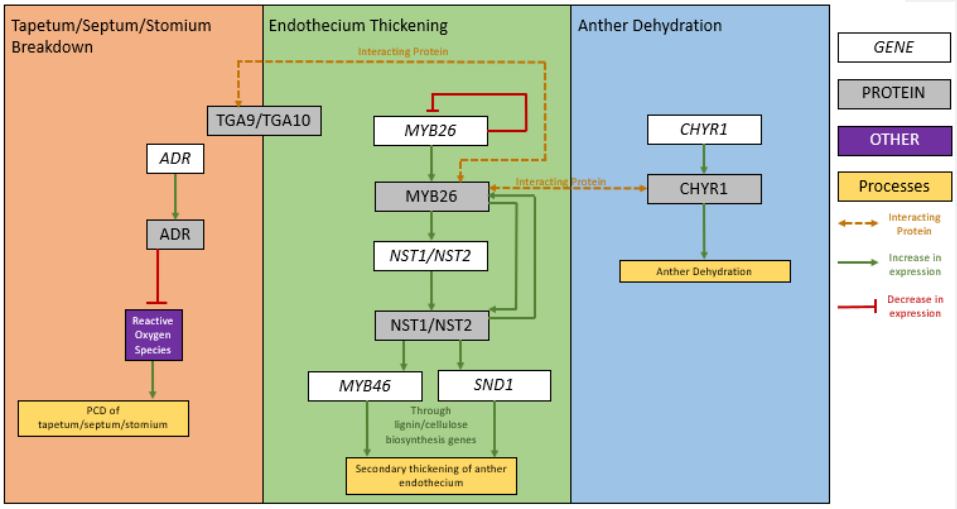

Figure 6.1 – A visual summary of previously suggested interactions involved in the *MYB26* driven, anther indehiscence network. It highlights *MYB26, NST1* and *NST2* interactions suggested by Yang *et al.* (2007; Yang *et al.* 2017), along with proteins identified as interacting with MYB26 by Mo (2017). *CHYR1* seems to play a role in anther dehydration (Hsu et al. 2014), and ADR has been suggested playing a role in programmed cell death as part of anther dehiscence (Dai et al. 2019).

Male sterility in plants is important to investigate because it has commercial applications in crop systems. Heterosis (the process of hybrid crops producing a higher yield than the parent lines) is a way to maximise yields (Duvick 1999; Lippman and Zamir 2007), which is important for global food security (Godfray *et al*. 2010; Pinstrup-Andersen 2009). One issue with heterosis is the need to prevent self-fertilisation to allow for the development of hybrid crops. Commercially, this can be achieved through emasculation of plants (which has issues with maintaining lines without time consuming fertility restoration systems) or through frequently unreliable chemical gametocides (Wilson and Zhang 2009). A lot of male sterile plants have been identified with non-viable pollen (Feng and Dickinson 2007; Ma 2005), but there are issues with recovering fertility in these lines. An interesting approach is to examine plants

which produce viable pollen but are fertility impaired due to pollen not being released due to anther indehiscence (Wilson *et al*. 2011).

MYB26 has been shown to be driver of a key factor in successful anther dehiscence – the selective secondary cell wall thickening of the anther endothecium. It appears to regulate *NST1/NST2* expression, which in turn leads to the activation of secondary cell wall biosynthesis genes (Yang *et al.* 2017; Yang *et al*. 2007). Further expansion and analysis of the MYB26 driven network for anther endothecium thickening is important for understanding of anther dehiscence in Arabidopsis. Understanding these processes may help to develop male sterile lines in the future for commercial use in agricultural processes. The previously understood network is highlighted in Figure 6.1.

Here network modelling was used to simulate different interactions between NST1, NST2 and MYB26 to test whether the interactions between MYB26 and NST1/NST2 are at the transcription/translation stage or whether it involves some post-translational modification. This was compared to observed expression changes in mutant lines to develop a more detailed understanding of the interaction between these genes (Chapter 3).

SAF1 was investigated to determine its role within the MYB26 driven, anther endothecium thickening network (Chapter 4) with the hypothesis being that SAF1 works to remove NST1/NST2 from the anther endothecium to prevent premature secondary thickening. It is also suggested that MYB26 removes SAF1 from the system to allow the accumulation of NST1/NST2. Other potential redundant genes and genes identified by Mo (2017) were examined to determine 1) if they are involved in anther endothecium thickening and 2) where they fit within the expanding network which is being developed (Chapter 5).

Finally, the mathematical simulated model (Chapter 3) is revisited to further determine hypothesised interactions with new genes within the network.

6.1 *NST2* Appears to Promote *MYB26* Expression by Slowing Removal of MYB26 Protein

Previous work (Yang *et al*. 2017; Yang *et al*. 2007) has suggested that anther endothecium secondary thickening is regulated by a key transcription factor *MYB26*, which drives secondary cell wall biosynthesis through upregulation of two NAC-domain transcription factors, *NST1* and *NST2* in anther tissues. NAC transcription factors are known to be important for secondary cell wall biosynthesis (McCahill and Hazen 2019), with genes such as *NST3* playing a very key role secondary cell wall synthesis of woody tissues in *Arabidopsis* (Mitsuda *et al.* 2007). *NST1* has a role in driving secondary cell wall biosynthesis in vascular tissues (Mitsuda *et al*. 2007; Zhong *et al*. 2007) along with being shown to be important for anther dehiscence (Mitsuda *et al*. 2005). *NST2* is also known to play a role in anther dehiscence along with expression being localised to developing anthers (Mitsuda *et al*. 2005). This, combined with the known role of NAC transcription factors in secondary cell development in other tissues strongly suggests that *NST1/NST2* play a role specifically in secondary cell wall biosynthesis in the anther endothecium, as previously proposed (Yang *et al*. 2017; Yang *et al*. 2007).

The relationship between these genes is not that *MYB26* simply drives *NST1*/*NST2* expression despite the presence of a MYB26 binding motif present in *NST1*/*NST2* promoter sequences (Mo 2017). This can be shown by the fact that overexpressing *NST1* or *NST2* does not recover phenotype in a *myb26* background (Yang *et al.* 2017), and the fact that *MYB26* did not bind with *NST1/NST2 in vivo* when analysed with ChIP analysis (Mo 2017).

Some of the interactions between *MYB26, NST1* and *NST2* have been deduced from the expression analysis by Yang *et al.* (2017), with *MYB26* found to directly drive *NST1/NST2* expression, but the lack of fertility recovery in *myb26* plants with induced *NST1* or *NST2* expression led to the conclusion that the removal

of NST1/NST2 proteins was more important (Yang *et al*. 2017). The work by Yang *et al.* (2017) also concluded that *NST2* leads to an upregulation of *MYB26* expression, whilst it appears *MYB26* is self-inhibitory. However, it was not clear from this work whether the upregulation of *MYB26* by *NST2* was due to *NST2* driving *MYB26* expression, or whether NST2 prevents MYB26 from being degraded (network interactions highlighted in Figure 3.3b or Figure 3.4b).

Using 2 different models with interactions of *NST2* and *MYB26* altered between the two, this work investigated which interaction (driving gene expression or slowing protein breakdown) mathematically led to simulated relative expression levels depending on the mutation which were most similar to the observed data by Yang et al. (2017). Two models were designed to investigate whether *NST2* drove *MYB26* transcription and translation (these were assumed to be directly proportional for ease of model coding, i.e. the ratio of DNA transcription and protein translation was assumed to be 1:1) (Figure 3.3), or whether it slowed protein breakdown of MYB26 (Figure 3.4). Model 1 (Section 3.3.1) seems to align better with the observed data from Yang et al. (2017) than model 2, suggesting that *MYB26* upregulation by *NST2* occurs at the post-translational level. The differences between simulated expressions with model 1 and model 2 when compared to the observed data (Figure 3.2), suggests that there is a more complication relationship between these 3 genes and anther endothecium secondary thickening, than just direct regulation via a linear path.

As previously discussed (Chapter 3) there are some limitations within these models. They assume transcription and translation are directly proportional. One has to assume that protein levels are zero until gene expression is activated. It is impossible to separate translation and transcription, and so if there is any interaction at the mRNA level that affects gene expression, then this would not be detected by these models. It may be useful to separate out the rate of change of expression equations into the rate of change of mRNA (upregulation through transcription of DNA minus downregulation by

breakdown) and rate of change of protein (upregulation by mRNA levels translation minus downregulation by protein breakdown) to better investigate very specific interactions. This is discussed further in Section 6.5.3.

6.2 *SAF1* could play a role in the removal of NST1/NST2

Overexpression of *NST1* or *NST2* did not recover the phenotype in *myb26* plants, but overexpression of both of these genes did lead to ectopic secondary cell wall deposition in the epidermal cells, however, not in the endothecium. This may be indicative of a repressor of NST1/NST2 being present in the developing anther endothecium. The relationship of MYB26 promoting NST1/NST2 expression by allowing protein accumulation, which was suggested by Yang *et al.* (2017), coupled with this observation, could suggest the secondary role of *MYB26* in *NST1/NST2* expression during anther endothecium thickening, is the removal of this hypothetical repressor.

One gene which was previously identified as playing a role in anther endothecium thickening is *SAF1* (Kim *et al*. 2012), which was investigated here with the hypothesis that it has a potential role in the *MYB26* network. Kim *et al*. (2012) identified *SAF1* as having a role in male fertility, with overexpression lines having male sterility. This was due to a lack of anther dehiscence because of an absence of anther endothecium thickening. This was putatively caused by a down regulation of secondary cell wall biosynthesis genes in the *SAF1* overexpression plants. Knowing that NAC transcription factors activate secondary cell wall biosynthesis genes, via MYB proteins such as *MYB46* and *MYB83* (Taylor-Teeples *et al*. 2015; Zhang *et al*. 2018b), it is possible that this lack of secondary cell wall thickening in the endothecium is due to a lack of NST1/NST2 or MYB26 in the anther endothecium.

SAF1 encodes an F-box protein (Kim *et al*. 2012) which are important in development by the removal of proteins within the cell through the process of ubiquitination (Nandi *et al*. 2006; Schumann *et al.* 2011). F-box proteins work

as part of a Skp1-Cullin-F-box (SCF), which is the most common E3 ligase in plants. E3 ligases generally add ubiquitin to target proteins (Schumann *et al.* 2011) which then leads to the removal of the protein by the 26S proteasome, although there is evidence that some E3 ligases work to remove proteins through the N-end rule pathway (Jin *et al.* 2004; Skaar *et al.* 2009). It is therefore likely that SAF1 functions to remove a protein(s) and given the observed phenotype of the overexpression lines that this protein target is part of the *MYB26* network to prevent anther endothecium thickening in overexpression lines.

The overexpression of *SAF1* leads to a downregulation of *MYB26* and all downstream genes (*NST1, NST2* and *IRX1*) (Figure 4.10). Using the knowledge that SAF1 is an F-box protein and likely works at a protein level then the likely protein being removed to lead to a downregulation in *MYB26* expression is NST2*,* since NST2 is known to promote *MYB26* expression (Yang *et al*. 2017). This is turn would lead to a downregulation of all downstream genes as is seen is Figure 4.10. Removal of MYB26 protein by SAF1 itself would be expected to lead to an upregulation of *MYB26* - this is due to the fact that MYB26 downregulates the expression of *MYB26* (Yang *et al.* 2017)and so less MYB26 within the system would likely lead to increased gene expression. The removal of an unknown protein upstream of *MYB26* would also lead to the results observed here. However if one assumes *MYB26* is the key driver of anther endothecium development (Yang et al. 2017), and based on work by Cecchetti *et al.* (2008, 2013) suggesting *MYB26* activation initiates in response to low auxin levels as auxin is removed from the endothecium, then *MYB26* may be assumed to be an initial activator gene for endothecium thickening in response to hormonal changes. Another supporting piece of evidence suggesting SAF1 interacts with NST2 is the lack of a phenotype outside of the anther in *35S::SAF1* plants. It's known that *NST2* is predominantly expressed in the anther and the removal of a protein with a less specific role would lead to

secondary cell wall issues throughout the plant, such as in vascular tissues. The assumption that SAF1 negatively regulates NST2 can be supported by the observation that overexpression *MYB26* leads to a knocking out of *SAF1* expression (Figure 4.11) whilst knocking out *MYB26* led to a high level (~20x) of *SAF1* upregulation. This, coupled with the suggestion that SAF1 works to inhibit anther endothecium thickening by removing NST2, allows one to hypothesise that *SAF1* fits into the *MYB26* network as shown in Figure 6.2.

A T-DNA insertion knock out mutant for *saf1* was investigated but it did not seem to affect *SAF1* expression. Further analysis was carried out for this SAIL T-DNA insert line and it was discovered that there was only partial transcript of *SAF1.* Looking at where the F-box motif is within the gene (Figure 5.6) it would appear that the partial transcription is unlikely to lead to a functional protein. The expression levels for *MYB26, NST1, NST2* and *IRX1* detected in SAIL_425_B06 (Figure 4.10) were not significantly different to wild type plants – the lack of a functional SAF1 protein in SAIL_425_B06 suggests the wild type gene expression in SAIL_425_B06 is down to *saf1* knock out not affecting the expression of these genes. This is supported by the network proposed network in Figure 6.2 as it would suggest that SAF1 prevents premature expression of *NST1* and *N10ST2. MYB26* turns on later in anther development (Steiner-Lange *et al*. 2003; Yang *et al*. 2007) when auxin levels are reduced (Cecchetti *et al*. 2013; Cecchetti *et al.* 2008) and in turn knocks down / knocks out *SAF1* to allow the downstream genes to be expressed. If this is the case, then knocking out *saf1* would not change the expression of *NST1/NST2/IRX1* at the appropriate time of anther development (stages 11-13), but may allow premature expression. This may lead to premature secondary cell wall thickening, however if one assumes that *MYB26* does directly drive *NST1/*NST2 transcription (along with its secondary role of removing SAF1) then *NST1/NST2* expression could still be temporally controlled. It could be the scenario where *SAF1* is a second level of control of ensuring anther endothecium thickening

occurs at the appropriate time in the appropriate tissues, due to the extremely well-co-ordinated nature of anther dehiscence processes in developing flowers. This was not examined here and is something that could be investigated in the future. One thing that should be noted is that the partial *SAF1* transcript does contain a FBD Domain (highlighted in Figure 5.6) which is a domain observed in F-Box proteins, but has an unknown function (Doerks *et al*. 2002) and therefore may have a role in the network.

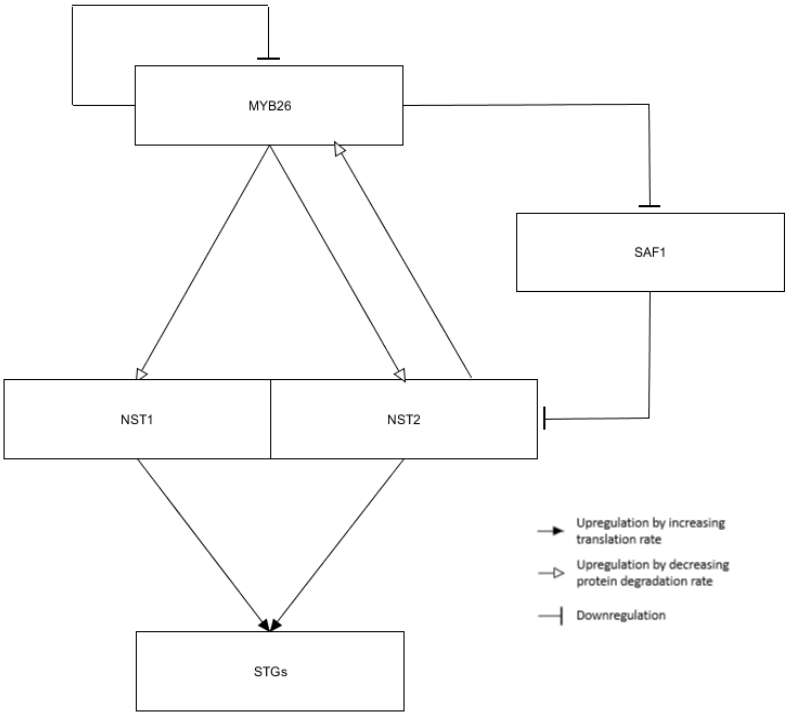

Figure 6.2 – Graphical representation of potential interactions within the anther endothecium secondary thickening gene network. Arrows with solid heads represent upregulation of gene expression (transcription and translation) whilst arrows with an open head represent upregulation of gene expression by slowing down the removal of the target gene's protein. A bar represents downregulation of gene expression however there is no distinction between method of downregulation.

6.3 *TGA9, TGA10* and *PKSP* appear to interact with the *MYB26* network

Anther endothecium secondary thickening occurs around stages 11 onwards (Goldberg *et al.* 1993; Smyth *et al.* 1990), coinciding with when *TGA9, TGA10* and *PKSP* expression increases (Figure 5.20 – Figure 5.23), which, coupled with knockout mutants for *tga9/tga10* having been shown to anther indehiscent (Mo 2017) due to a lack of endothecium thickening, suggests that these genes could be involved in anther endothecium thickening. Knockout *pksp* plants did not have a phenotype however it is very possible that this is due to the redundency with its putative orthologue *PCRK1* (*Kong et al.* 2016; Sreekanta *et al.* 2015).

PCRK1 expression is higher in developing leaves (Figure 5.23) and it therefore seems likely that it may play a role outside of anther development, although this does not exclude it from also playing a role in anther development alongside functions within other tissues.

TGA9 and *PKSP* expression is significantly downregulated when *SAF1* is overexpressed (Figure 5.24). Going back to the initial hypothesis that SAF1 removes NST2 from the system, leading to a reduction in *MYB26* expression, this could suggest that these two genes are positively regulated by *MYB26*, which was also suggested by Mo (2017), since *TGA10* and *PCRK1* both have a slightly reduced expression in the *35S::SAF1* line. This smaller change in relative expression could be explained by the fact that these genes are less anther specific (Figure 5.21 and Figure 5.23) and these two genes may be less important to anther dehiscence.

The *tga9/tga10* knock out lines used by Mo (2017) were investigated for relative expression. Relative expression of *SAF1* increased when *tga9/tga10* were knocked out (Figure 5.25), similar to with *myb26* plants (Figure 4.11). This may suggest that *SAF1* is downstream of *TGA9/TGA10,* possibly by *MYB26* activating *TGA9/TGA10* expression which in turn inhibits expression of *SAF1.* However, overexpression of *MYB26* did not lead to an increase in *TGA9* or *TGA10* expression (Figure 5.25) which would suggest that this is not the case. Mo (2017) initially identified TGA9 as a protein which may interact with MYB26, and so there is the possibility that *TGA9* is regulated by factors external from the *MYB26* network and works with MYB26 to inhibit *SAF1* expression, potentially through dimerisation. However, a further complication to this arises since *myb26* mutants have no expression of *TGA9* and *TGA10*, which supports the hypothesis that *MYB26,* directly or indirectly, upregulates their expression. *PKSP* expression is reduced when either *MYB26* or *NST1* and *NST2* are knocked out (Figure 5.24). Mo (2017) suggested that *PKSP* was downstream of *MYB26* as a direct target, which is supported by the reduced expression in the *myb26* line. However, overexpression of *MYB26* did not increase *PKSP* expression.

However further qRT-PCR experiments are needed to validate this. Reduced expression in the *nst1/nst2* double knockout plants suggests that *PKSP* expression is also upregulated by *NST1* or *NST2*, however it is as likely that the lower *PKSP* expression is a result of lower *MYB26* expression when *NST2* is knocked out due to the upregulation of *MYB26* by NST2 (Yang *et al*. 2017).

6.4 Summary

Anther dehiscence is a key aspect of male fertility in developing flowers (Goldberg *et al*. 1993; Sanders *et al*. 1999). A number of highly coordinated events occur to allow for the successful development and release of viable pollen, and the breakdown at any of these stages could lead to infertile plants. Understanding male fertility is important because it will allow for the potential development of hybrid crops easier which should improve yields (Duvick 1999).

Here a part of anther dehiscence has been investigated, looking at the specifics of a network which seems to be the key regulation in anther endothecium secondary thickening. This secondary thickening has been suggested to be driven by *MYB26,* which in turn activates *NST1/NST2*, which initiate secondary cell wall biosynthesis genes. The interaction between *MYB26, NST1* and *NST2,* was fairly well established, however it was not a direct activation of the NAC domain genes by *MYB26.* Additionally, mathematical modelling is used to suggest *NST2* drives MYB26 production by preventing protein breakdown. An extra F-box encoding gene *SAF1* is placed into the network with the role of removing NST2 from the cell. *MYB26* inhibits the expression of this gene. SAF1 may act as a temporal control of *NST2* accumulation in the anther. A second gene *TGA9* seems to have a role similar to *MYB26* in inhibiting *SAF1* expression, although it is unclear whether it works downstream of *MYB26* or alongside/with MYB26. Some potentially redundant genes have also been identified.

It is accepted that there is a high level of regulation in anther dehiscence during development due to the need for high levels of coordination (Zhong *et al*. 2008). *SAF1* acting to prevent premature expression of *NST2* is entirely plausible. The

whole network could be controlled by hormone levels to ensure that the auxin minima (Sorefan *et al.* 2009) in the endothecium coincides with an auxin maxima in other tissues. This maxima in other tissues allows for the breakdown of the stomium and septum (Cecchetti *et al.* 2017; Wilson *et al.* 2011), through both programmed cell death (Keijzer 1987) and reactive oxygen species (Goto-Yamada et al. 2014). Whilst this is occurring, an auxin minima within the endothecium would allow the expression of *MYB26* (Cecchetti *et al*. 2013; Cecchetti *et al*. 2008)*,* potentially linked to *ADR* expression being reduced, which has been shown to have an effect on *NST1/NST2* expression. The activation of *MYB26* drives *NST1/NST2* expression, although (according to the hypothesis suggested here) importantly also inhibits SAF1 expression (which is reduced in late stage anthers (Figure 5.12)) to allow the accumulation of NST2 (and NST1). These genes then turn on a suite of secondary cell wall biosynthesis genes which allow for secondary cell wall thickening to occur in the anther endothecium (McCahill and Hazen 2019; Mitsuda *et al.* 2005). The anther then dehydrates due to a number of factors which increase the osmotic pressure outside of the endothecium and epidermal cells, causing an increase in water transport from these, leading these cells to dehydrate (Bots *et al.* 2005; Matsui *et al*. 2000; Stadler *et al*. 1999). The differential forces exerted across these tissues due to the selective anther endothecium thickening and the breakdown

of other tissues then allows anther dehiscence to occur (Nelson *et al*. 2012). The overall network is summarised in Figure 6.3

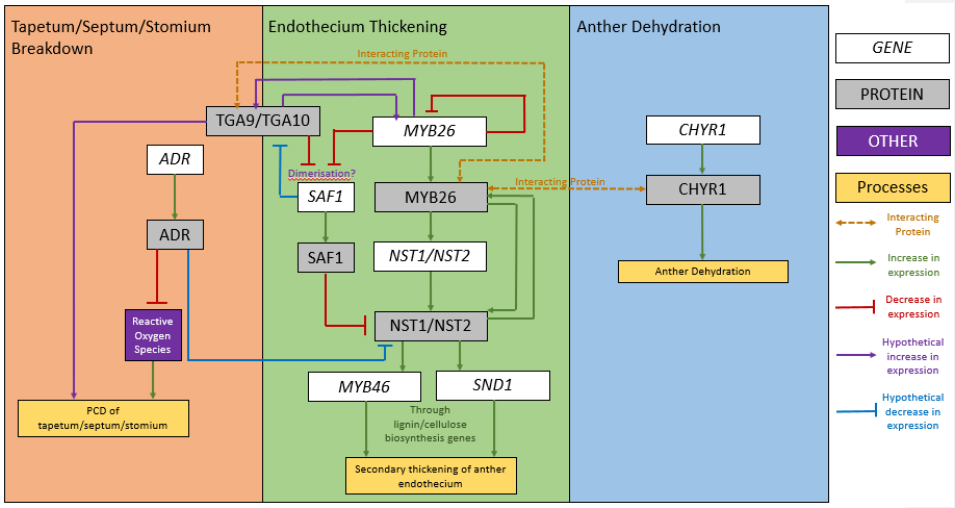

Figure 6.3 – A summary of the known genetic network interactions from before this thesis, along with *SAF1* and it's role within the network based on work carried out here. Additionally, the interactions of how the different aspects of anther dehiscence are synchronised and interlinked are also hypothesised in the figure.

6.5 Future Work

6.5.1 Transformation Repeats and Alternative Methods

Unfortunately there were issues whilst growing up plants carrying the NST2::GFP translational fusions. Seeds were sown on a *Hygromycin* selective plate for and germinated. However, growth arrested after approximately a week. The *Agrobacterium* were tested by colony PCRs and sequencing for the presence of the *NST2* plasmid. The arrest of growth may be a result of a low transformation efficiency due to plant health. Repeats of these transformations with the same plasmids (which have been sequenced completely) should result in successful transformations. Repeating transformations of *35S::SAF1* plants along with wild type using the ProNST2::NST2::GFP vector would be very interesting. It would be very important to visualise NST2 localisation in *SAF1* overexpression plants – if the network highlighted in Figure 6.2 is correct then

NST2 should be visible in wild type plants, but reduced or removed in *SAF1* overexpression plants. This could be strong evidence that SAF1 is involved in the *MYB26* genetic network by negatively regulating NST2 accumulation. In addition to this, it would be useful to also repeat the CRISPR transformation of wild type plants to generate additional *saf1* mutant lines to investigate whether the SAIL_426_B06 plants did successfully knock out the *SAF1* expression or whether a different method to knock the gene out (CRISPR) would provide different information.

6.5.2 Further Investigation of Other Genes

The evidence presented suggest that *TGA9/TGA10* could be upstream of *SAF1* and play a role in the negative regulation of *SAF1* expression. However, it is not clear where it fits in the network with regards to *MYB26, NST1/NST2*, with there being contrasting evidence suggesting that it may/may not be a direct target of *MYB26.* Repeats of *TGA9/TGA10* expression in *myb26* and *35S::MYB26* lines could shed some light on this. Previous work (Wilson Lab, unpublished) showed that overexpression of *TGA9* led to an increase in *MYB26* expression – this could be due to an increase in NST2 accumulation (if *SAF1* expression in inhibited) and transforming this line with ProNST2::NST2::GFP construct could help clarify this. Analysis of how overexpression of *TGA9* affects the expression of *NST1/NST2* would also be interesting to explore, although if one assumes that *TGA9* inhibits *SAF1* expression then it would not be clear whether *TGA9* was changing *NST1/NST2* expression via this. To try and understand if *TGA9* was more directly responsible for *NST1/NST2* activation, then it would be interesting to investigate protein-DNA interactions, for example with ChIP assay analysis, or alternatively investigate protein-protein interactions with luciferase assays or yeast-2-hybrid analysis.

PKSP is downregulated when *SAF1* is overexpressed. It is suggested that it is downstream as a direct target of *MYB26* (Mo 2017), and *SAF1* overexpression leads to a downregulation of *MYB26* expression. *PKSP* was identified using ChIP

assays of MYB26 putative targets and so should be directly downstream of *MYB26,* although it is not clear what role *PKSP* plays. Since *pksp* plants had no phenotype (Mo 2017), it would be worthwhile to generate *pksp/pcrk1* mutants to see if these had a visible phenotype. This mutant could also be used to see if knocking out *PKSP* affects expression of *NST1/NST2* or *IRX1.* The same analysis should be carried out on an overexpression line for *PKSP.* Finally, it is not clear whether *PKSP* plays a role anther endothecium thickening and mutant analysis for phenotypes along with analysis to determine anything downstream of *PKSP* would be especially useful to define its function.

6.5.3 Development of future models

One thing which is important to do in the future would be to try and develop network models including *SAF1, TGA9/TGA10* and *PKSP/PCRK1.* Looking at the network illustrated in Figure 6.2, the potential role of *SAF1* is hypothesised to have quite a specific role. It is inhibited by *MYB26* at the transcription level, and in turn inhibits NST2 accumulation at the post-translation level (by ubiquitination of NST2). To insert this into the network then equations should be altered as follows, with a new equation for the expression of *SAF1.* Green within the equation indicates a new section/new equation.

$$\frac{d(NST1_{mRNA})}{dt} = \frac{\alpha_{NST1_{mRNA}}}{\frac{\eta \cdot SAF1}{K_{S3}}} - \frac{K_{NST1_{mRNA}} \cdot \beta_{NST1_{mRNA}} \cdot NST1_{mRNA}}{K_{NST1_{mRNA}} + MYB26}$$

$$\frac{d(NST1_{protein})}{dt} = \sigma \cdot \varepsilon_{NST1_{mRNA}} \cdot NST1_{mRNA} - \tau \cdot \theta_{NST1_{protein}} \cdot SAF1_{protein} \cdot NST1_{protein}$$

$$\frac{d(NST2_{mRNA})}{dt} = \frac{\alpha_{NST2_{mRNA}}}{\frac{\eta \cdot SAF1}{K_{S3}}} - \frac{K_{NST2_{mRNA}} \cdot \beta_{NST2_{mRNA}} \cdot NST2_{mRNA}}{K_{NST2_{mRNA}} + MYB26_{protein}}$$

$$\frac{d(NST2_{protein})}{dt} = \lambda \cdot \varepsilon_{NST2_{mRNA}} \cdot NST2_{mRNA} - \nu \cdot \theta_{NST2_{protein}} \cdot SAF1_{protein} \cdot NST2_{protein}$$

$$\frac{d(MYB26_{mRNA})}{dt} = \alpha_{MYB26_{mRNA}} \cdot \frac{c + \frac{MYB26_{mRNA}}{K_{M1}}}{1 + \frac{NST2_{protein}}{K_{M1}} + \frac{\phi \cdot MYB26_{mRNA}}{\delta}} - \beta_{MYB26_{mRNA}} \cdot MYB26_{mRNA}$$

$$\frac{d(MYB26_{protein})}{dt} = \varphi \cdot \varepsilon_{MYB26_{mRNA}} \cdot MYB26_{mRNA} - \psi \cdot \theta_{MYB26_{protein}} \cdot MYB26_{protein}$$

$$\frac{d(SAF1_{mRNA})}{dt} = \frac{\alpha_{SAF_{mRNA}}}{\phi \cdot MYB26_{protein}} - \beta_{STG} \cdot SAF_{mRNA}$$

$$\frac{d(SAF1_{protein})}{dt} = \vartheta \cdot \varepsilon_{SAF1_{mRNA}} \cdot SAF1_{mRNA} - \varkappa \cdot \theta_{SAF1_{protein}} \cdot MYB26_{protein} \cdot SAF1_{protein}$$

Coding these new equations into the Matlab models and simulating relative gene expression would be interested to see if this could get simulated expressions to correspond to the observed data.

It would also be interesting to try and develop equations for *TGA9, TGA10,* and *PKSP,* however without further analysis of these genes it is impossible to conclude where exactly within the network these genes might feature.

References

Alexander M (1969) Differential staining of aborted and nonaborted pollen. Stain technology 44 (3):117-122

Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, Elledge SJ (1996) SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell 86 (2):263-274

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9 (1):39

Berendzen K, Searle I, Ravenscroft D, Koncz C, Batschauer A, Coupland G, Somssich IE, Ülker B (2005) A rapid and versatile combined DNA/RNA extraction protocol and its application to the analysis of a novel DNA marker set polymorphic between *Arabidopsis thaliana* ecotypes Col-0 and *Landsberg erecta*. Plant Methods 1 (1):4

Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annual Review of Genetics 45:273-297

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326 (5959):1509-1512

Bonner LJ, Dickinson H (1990) Anther dehiscence in *Lycopersicon esculentum*: II. Water relations. New Phytologist 115 (2):367-375

Bots M, Feron R, Uehlein N, Weterings K, Kaldenhoff R, Mariani T (2004) *PIP1* and *PIP2* aquaporins are differentially expressed during tobacco anther and stigma development. Journal of Experimental Botany 56 (409):113-121

Bots M, Vergeldt F, Wolters-Arts M, Weterings K, van As H, Mariani C (2005) Aquaporins of the *PIP2* class are required for efficient anther dehiscence in tobacco. Plant Physiology 137 (3):1049-1056

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. The Plant Cell 17 (8):2281-2295

Browne RG, Iacuone S, Li SF, Dolferus R, Parish RW (2018) Anther morphological development and stage determination in *Triticum aestivum*. Frontiers in Plant Science 9:228

Cardozo T, Pagano M (2004) The SCF ubiquitin ligase: insights into a molecular machine. Nature Reviews Molecular Cell Biology 5 (9):739

Cecchetti V, Altamura MM, Brunetti P, Petrocelli V, Falasca G, Ljung K, Costantino P, Cardarelli M (2013) Auxin controls *Arabidopsis* anther dehiscence by regulating endothecium lignification and jasmonic acid biosynthesis. The Plant Journal 74 (3):411-422

Cecchetti V, Altamura MM, Falasca G, Costantino P, Cardarelli M (2008) Auxin regulates Arabidopsis anther dehiscence, pollen maturation, and filament elongation. The Plant Cell 20 (7):1760-1774

Cecchetti V, Celebrin D, Napoli N, Ghelli R, Brunetti P, Costantino P, Cardarelli M (2017) An auxin maximum in the middle layer controls stamen development and pollen maturation in Arabidopsis. New Phytologist 213 (3):1194-1207

Cecchetti V, Pomponi M, Altamura MM, Pezzotti M, Marsilio S, D'Angeli S, Tornielli GB, Costantino P, Cardarelli M (2004) Expression of rolB in tobacco flowers affects the coordinated processes of anther dehiscence and style elongation. The Plant Journal 38 (3):512-525

Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, Collingwood TN, Frodin M, Davis GD (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nature methods 8 (9):753

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. The Plant Journal 89 (4):789-804

Cheng S-H, Zhuang J-Y, Fan Y-Y, Du J-H, Cao L-Y (2007) Progress in research and development on hybrid rice: a super-domesticate in China. Annals of botany 100 (5):959-966

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186 (2):757-761

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal 16 (6):735-743

Coakley SM, Scherm H, Chakraborty S (1999) Climate change and plant disease management. Annual Review of Phytopathology 37 (1):399-426

Cooper G, Hausman R (2000) The cell: a molecular approach. Sinauer Associates. Sunderland, MA

Dai S-Y, Hsu W-H, Yang C-H (2019) The Gene ANTHER DEHISCENCE REPRESSOR (ADR) Controls Male Fertility by Suppressing the ROS Accumulation and Anther Cell Wall Thickening in *Arabidopsis*. Scientific Reports 9 (1):5112

Dawson J, Sözen E, Vizir I, Van Waeyenberge S, Wilson Z, Mulligan B (1999) Characterization and genetic mapping of a mutation (*ms35*) which prevents anther dehiscence in *Arabidopsis thaliana* by affecting secondary wall thickening in the endothecium. The New Phytologist 144 (2):213-222

Dawson J, Wilson ZA, Aarts MGM, Braithwaite AF, Briarty LG, Mulligan BJ (1993) Microspore and pollen development in six male-sterile mutants of *Arabidopsis thaliana*. Canadian Journal of Botany 71 (4):629-638. doi:10.1139/b93-072

Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, Naylor RL (2018) Increase in crop losses to insect pests in a warming climate. Science 361 (6405):916-919

Dill A, Thomas SG, Hu J, Steber CM, Sun T-p (2004) The *Arabidopsis* F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. The Plant Cell 16 (6):1392-1405

Doerks T, Copley RR, Schultz J, Ponting CP, Bork P (2002) Systematic identification of novel protein domain families associated with nuclear functions. Genome Research 12 (1):47-56

Dong J, Kim ST, Lord EM (2005) Plantacyanin plays a role in reproduction in *Arabidopsis*. Plant Physiology 138 (2):778-789

Duvick DN (1999) Heterosis: feeding people and protecting natural resources. The Genetics and Exploitation of Heterosis in Crops:19-29

Elledge SJ, Harper JW (1998) The role of protein stability in the cell cycle and cancer. Biochimica et Biophysica Acta 1377 (2):M61-70

FAO I (2014) WFP. strengthening the enabling environment for food security and nutrition. Rome: FAO:2014

Fauser F, Roth N, Pacher M, Ilg G, Sánchez-Fernández R, Biesgen C, Puchta H (2012) In planta gene targeting. Proceedings of the National Academy of Sciences 109 (19):7535-7540

Feldmann KA (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. The Plant Journal 1 (1):71-82

Feng X, Dickinson HG (2007) Packaging the male germline in plants. Trends in Genetics 23 (10):503-510

Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468 (7320):67

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010) Food Security: The Challenge of Feeding 9 billion People. science 327 (5967):812-818

Goldberg RB, Beals TP, Sanders PM (1993) Anther development: basic principles and practical applications. The Plant Cell 5 (10):1217

Gómez JF, Talle B, Wilson ZA (2015) Anther and pollen development: a conserved developmental pathway. Journal of Integrative Plant Biology 57 (11):876-891

González-Carranza ZH, Rompa U, Peters JL, Bhatt AM, Wagstaff C, Stead AD, Roberts JA (2007) *HAWAIIAN SKIRT*: an F-box gene that regulates organ fusion and growth in *Arabidopsis*. Plant Physiology 144 (3):1370-1382

Goto-Yamada S, Mano S, Nishimura M (2014) The role of peroxisomes in plant reproductive processes. In: Sexual Reproduction in Animals and Plants. Springer, Tokyo, pp 419-429

Gregory PJ, Ingram JS, Brklacich M (2005) Climate change and food security. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1463):2139-2148

Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover III CV, Graveley BR, Terns RM (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Molecular cell 45 (3):292-302

Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327 (5962):167-170

Hsu K-H, Liu C-C, Wu S-J, Kuo Y-Y, Lu C-A, Wu C-R, Lian P-J, Hong C-Y, Ke Y-T, Huang J-H (2014) Expression of a gene encoding a rice RING zinc-finger protein, OsRZFP34, enhances stomata opening. Plant Molecular Biology 86 (1-2):125-137

Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA (2005) FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. Science 309 (5732):293-297

Ishiguro S, Kawai-Oda A, Ueda J, Nishida I, Okada K (2001) The DEFECTIVE IN ANTHER DEHISCENCE1 gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in *Arabidopsis*. The Plant Cell 13 (10):2191-2209

Jansen R, Embden JDv, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology 43 (6):1565-1575

Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW (2004) Systematic analysis and nomenclature of mammalian F-box proteins. Genes & Development 18 (21):2573-2580

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337 (6096):816-821

Jones L, Ennos AR, Turner SR (2001) Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of *Arabidopsis*. The Plant Journal 26 (2):205-216

Kawanabe T, Ariizumi T, Kawai-Yamada M, Uchimiya H, Toriyama K (2006) Abolition of the tapetum suicide program ruins microsporogenesis. Plant and Cell Physiology 47 (6):784-787

Keijzer C (1987) The processes of anther dehiscence and pollen dispersal: I. The opening mechanism of longitudinally dehiscing anthers. New Phytologist 105 (3):487-498

Kepinski S, Leyser O (2005) The *Arabidopsis* F-box protein TIR1 is an auxin receptor. Nature 435 (7041):446

Kim HS, Delaney TP (2002) *Arabidopsis* SON1 is an F-box protein that regulates a novel induced defense response independent of both salicylic acid and systemic acquired resistance. The Plant Cell 14 (7):1469-1482

Kim O-K, Jung J-H, Park C-M (2010) An *Arabidopsis* F-box protein regulates tapetum degeneration and pollen maturation during anther development. Planta 232 (2):353-366

Kim YY, Jung KW, Jeung JU, Shin JS (2012) A novel F-box protein represses endothecial secondary wall thickening for anther dehiscence in *Arabidopsis thaliana*. Journal of plant physiology 169 (2):212-216

Kirby E (2002) Botany of the wheat plant. Bread Wheat: Improvements and Production

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA (2016) A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. The Plant Journal 88 (6):1058-1070

Kong Q, Sun T, Qu N, Ma J, Li M, Cheng Y-t, Zhang Q, Wu D, Zhang Z, Zhang Y (2016) Two redundant receptor-like cytoplasmic kinases function downstream of pattern recognition receptors to regulate activation of SA biosynthesis. Plant physiology 171 (2):1344-1354

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in Arabidopsis. The plant cell 11 (12):2283-2290

Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome biology 8 (4):R61

Lehti-Shiu MD, Shiu S-H (2012) Diversity, classification and function of the plant protein kinase superfamily. Philosophical Transactions of the Royal Society B: Biological Sciences 367 (1602):2619-2639

Lehti-Shiu MD, Zou C, Hanada K, Shiu S-H (2009) Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. Plant Physiology 150 (1):12-26

Li N, Zhang D-S, Liu H-S, Yin C-S, Li X-x, Liang W-q, Yuan Z, Xu B, Chu H-W, Wang J (2006) The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. The Plant Cell 18 (11):2999-3014

Lin W, Ma X, Shan L, He P (2013) Big roles of small kinases: The complex functions of receptor-like cytoplasmic kinases in plant immunity and development. Journal of Integrative Plant Biology 55 (12):1188-1197

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. Trends in Genetics 23 (2):60-66

Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Biol 56:393-434

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF (2011) Evolution and classification of the CRISPR–Cas systems. Nature Reviews Microbiology 9 (6):467

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proceedings of the National Academy of Sciences 100 (21):11980-11985

Mariani C, De Beuckeleer M, Truettner J, Leemans J, Goldberg RB (1990) Induction of male sterility in plants by a chimaeric ribonuclease gene. Nature 347 (6295):737

Matsui T, Omasa K, Horie T (2000) High temperature at flowering inhibits swelling of pollen grains, a driving force for thecae dehiscence in rice (Oryza sativa L.). Plant Production Science 3 (4):430-434

McCahill IW, Hazen SP (2019) Regulation of Cell Wall Thickening by a Medley of Mechanisms. Trends in plant science

Miki D, Zhu P, Zhang W, Mao Y, Feng Z, Huang H, Zhang H, Li Y, Liu R, Zhang H (2017) Efficient generation of diRNAs requires components in the posttranscriptional gene silencing pathway. Scientific reports 7 (1):301

Miller JC, Holmes MC, Wang J, Guschin DY, Lee Y-L, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology 25 (7):778

Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M (2007) NAC transcription factors, *NST1* and *NST3*, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. The Plant Cell 19 (1):270-280

Mitsuda N, Ohme-Takagi M (2008) NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. The Plant Journal 56 (5):768-778

Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M (2005) The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. The Plant Cell 17 (11):2993-3006

Mo R (2017) Characterising the regulatory network of MYB26 during anther dehiscence. University of Nottingham, University of Nottingham

Murmu J, Bush MJ, DeLong C, Li S, Xu M, Khan M, Malcolmson C, Fobert PR, Zachgo S, Hepworth SR (2010) Arabidopsis basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. Plant Physiology 154 (3):1492-1504

Nandi D, Tahiliani P, Kumar A, Chandu D (2006) The Ubiquitin-Proteasome System. Journal of Biosciences 31 (1):137-155

Nelson M, Band L, Dyson R, Lessinnes T, Wells D, Yang C, Everitt N, Jensen O, Wilson Z (2012) A biomechanical model of anther opening reveals the roles of dehydration and secondary thickening. New Phytologist 196 (4):1030-1037

Nemhauser JL, Feldman LJ, Zambryski PC (2000) Auxin and ETTIN in Arabidopsis gynoecium morphogenesis. Development 127 (18):3877-3888

Okada K, Ueda J, Komaki MK, Bell CJ, Shimura Y (1991) Requirement of the auxin polar transport system in early stages of Arabidopsis floral bud formation. The Plant Cell 3 (7):677-684

Olsen AN, Ernst HA, Leggio LL, Skriver K (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends in Plant Science 10 (2):79-87

Parish RW, Li SF (2010) Death of a tapetum: a programme of developmental altruism. Plant Science 178 (2):73-89

Park JH, Halitschke R, Kim HB, Baldwin IT, Feldmann KA, Feyereisen R (2002) A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in *Arabidopsis* due to a block in jasmonic acid biosynthesis. The Plant Journal 31 (1):1-12

Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee Y-L (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nature Biotechnology 26 (7):808

Pickart CM (2001) Mechanisms underlying ubiquitination. Annual Review of Biochemistry 70 (1):503-533

Piffanelli P, Ross JH, Murphy D (1998) Biogenesis and function of the lipidic structures of pollen grains. Sexual Plant Reproduction 11 (2):65-80

Pinstrup-Andersen P (2009) Food security: definition and measurement. Food Security 1 (1):5-7

Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, Genschik P (2003) EIN3-dependent regulation of plant ethylene hormone signaling by two *Arabidopsis* F box proteins: EBF1 and EBF2. Cell 115 (6):679-689

Prigge MJ, Greenham K, Zhang Y, Santner A, Castillejo C, Mutka AM, O'Malley RC, Ecker JR, Kunkel BN, Estelle M (2016) The *Arabidopsis* auxin receptor F-box proteins AFB4 and AFB5 are required for response to the synthetic auxin picloram. G3: Genes, Genomes, Genetics 6 (5):1383-1390

Pul Ü, Wurm R, Arslan Z, Geißen R, Hofmann N, Wagner R (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Molecular Microbiology 75 (6):1495-1512

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nature Protocols 8 (11):2281

Rehman S, Yun S (2006) Developmental regulation of K accumulation in pollen, anthers, and papillae: are anther dehiscence, papillae hydration, and pollen swelling leading to pollination and fertilization in barley (*Hordeum vulgare L.*) regulated by changes in K concentration? Journal of Experimental Botany 57 (6):1315-1321

Reis A, Hornblower B, Robb B, Tzertzinis G (2014) CRISPR/Cas9 and targeted genome editing: a new era in molecular biology. NEB expressions 1:3-6

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiology 124 (2):495-498

Saleh-Gohari N, Helleday T (2004) Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. Nucleic acids research 32 (12):3683-3688

Salehin M, Bagchi R, Estelle M (2015) SCFTIR1/AFB-based auxin perception: mechanism and role in plant growth and development. The Plant Cell 27 (1):9-19

Sanders PM, Bui AQ, Weterings K, McIntire K, Hsu Y-C, Lee PY, Truong MT, Beals T, Goldberg R (1999) Anther developmental defects in *Arabidopsis* thaliana male-sterile mutants. Sexual Plant Reproduction 11 (6):297-322

Sanders PM, Lee PY, Biesgen C, Boone JD, Beals TP, Weiler EW, Goldberg RB (2000) The Arabidopsis DELAYED DEHISCENCE1 gene encodes an enzyme in the jasmonic acid synthesis pathway. The Plant Cell 12 (7):1041-1061

Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, Zhang F (2012) A transcription activator-like effector toolbox for genome engineering. Nature Protocols 7 (1):171

Santner A, Estelle M (2010) The ubiquitin-proteasome system regulates plant hormone signaling. The Plant Journal 61 (6):1029-1040

Schumann N, Navarro-Quezada A, Ullrich K, Kuhl C, Quint M (2011) Molecular evolution and selection patterns of plant F-box proteins with C-terminal kelch repeats. Plant physiology 155 (2):835-850

Scott RJ, Spielman M, Dickinson HG (2004) Stamen Structure and Function. The Plant Cell 16 (suppl 1):S46-S60. doi:10.1105/tpc.017012

Shen H, Strunks GD, Klemann BJ, Hooykaas PJ, de Pater S (2017) CRISPR/Cas9-induced double-strand break repair in *Arabidopsis* nonhomologous end-joining mutants. G3: Genes, Genomes, Genetics 7 (1):193-202

Shi J, Tan H, Yu X-H, Liu Y, Liang W, Ranathunge K, Franke RB, Schreiber L, Wang Y, Kai G (2011) Defective pollen wall is required for anther and microspore development in rice and encodes a fatty acyl carrier protein reductase. The Plant Cell 23 (6):2225-2246

Shiu S-H, Bleecker AB (2001) Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. Proceedings of the National Academy of Sciences 98 (19):10763-10768

Skaar JR, D'Angiolella V, Pagan JK, Pagano M (2009) SnapShot: F box proteins II. Cell 137 (7):1358. e1351-1358. e1352

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in *Arabidopsis*. The Plant Cell 2 (8):755-767

Sorefan K, Girin T, Liljegren SJ, Ljung K, Robles P, Galván-Ampudia CS, Offringa R, Friml J, Yanofsky MF, Østergaard L (2009) A regulated auxin minimum is required for seed dispersal in *Arabidopsis*. Nature 459 (7246):583

Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. Annual Review of Biochemistry 82:237-266

Sreekanta S, Bethke G, Hatsugai N, Tsuda K, Thao A, Wang L, Katagiri F, Glazebrook J (2015) The receptor-like cytoplasmic kinase PCRK 1 contributes to pattern-triggered immunity against Pseudomonas syringae in *Arabidopsis thaliana*. New Phytologist 207 (1):78-90

Stadler R, Truernit E, Gahrtz M, Sauer N (1999) The AtSUC1 sucrose carrier may represent the osmotic driving force for anther dehiscence and pollen tube growth in Arabidopsis. The Plant Journal 19 (3):269-278

Steiner-Lange S, Unte US, Eckstein L, Yang C, Wilson ZA, Schmelzer E, Dekker K, Saedler H (2003) Disruption of *Arabidopsis thaliana* MYB26 results in male sterility due to non-dehiscent anthers. The Plant Journal 34 (4):519-528

Stern KR, Bidlack JE, Jansky SH (2011) Stern's Introductory Plant Biology. McGraw-Hill,

Stintzi A (2000) The *Arabidopsis* male-sterile mutant, opr3, lacks the 12-oxophytodienoic acid reductase required for jasmonate synthesis. Proceedings of the National Academy of Sciences 97 (19):10625-10630

Stone JM, Walker JC (1995) Plant protein kinase families and signal transduction. Plant Physiology 108 (2):451-457

Stone SL, Callis J (2007) Ubiquitin ligases mediate growth and development by promoting protein death. Current Opinion in Plant Biology 10 (6):624-632

Taylor-Teeples M, Lin L, De Lucas M, Turco G, Toal T, Gaudinier A, Young N, Trabucco G, Veling M, Lamothe R (2015) An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. Nature 517 (7536):571

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. The Plant Cell 12 (12):2529-2539

Taylor NG, Scheible W-R, Cutler S, Somerville CR, Turner SR (1999) The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. The Plant Cell 11 (5):769-779

Thévenin J, Pollet B, Letarnec B, Saulnier L, Gissot L, Maia-Grondard A, Lapierre C, Jouanin L (2011) The simultaneous repression of CCR and CAD, two enzymes of the lignin biosynthetic pathway, results in sterility and dwarfism in *Arabidopsis thaliana*. Molecular Plant 4 (1):70-82

Varnier A-L, Mazeyrat-Gourbeyre F, Sangwan RS, Clément C (2005) Programmed cell death progressively models the development of anther sporophytic tissues from the tapetum and is triggered in pollen grains during maturation. Journal of Structural Biology 152 (2):118-128

Vierstra RD (2009) The ubiquitin–26S proteasome system at the nexus of plant biology. Nature Reviews Molecular Cell Biology 10 (6):385

Vij S, Giri J, Dansana PK, Kapoor S, Tyagi AK (2008) The receptor-like cytoplasmic kinase (OsRLCK) gene family in rice: organization, phylogenetic relationship, and expression during development and stress. Molecular Plant 1 (5):732-750

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153 (4):910-918

Wang Z-P, Xing H-L, Dong L, Zhang H-Y, Han C-Y, Wang X-C, Chen Q-J (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates

homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. Genome Biology 16 (1):144

Weissman AM (2001) Ubiquitin and proteasomes: themes and variations on ubiquitylation. Nature reviews Molecular Cell Biology 2 (3):169

Wilson ZA, Song J, Taylor B, Yang C (2011) The final split: the regulation of anther dehiscence. Journal of Experimental Botany 62 (5):1633-1649

Wilson ZA, Zhang D-B (2009) From *Arabidopsis* to rice: pathways in pollen development. Journal of experimental botany 60 (5):1479-1492

Wood AJ, Lo T-W, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X (2011) Targeted genome editing across species using ZFNs and TALENs. Science 333 (6040):307-307

Wu H-m, Cheung AY (2000) Programmed cell death in plant reproduction. In: Programmed Cell Death in Higher Plants. Springer, pp 23-37

Xu FX, Chye ML (1999) Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. The Plant Journal 17 (3):321-327

Xu L, Liu F, Lechner E, Genschik P, Crosby WL, Ma H, Peng W, Huang D, Xie D (2002) The SCFCOI1 ubiquitin-ligase complexes are required for jasmonate response in *Arabidopsis*. The Plant Cell 14 (8):1919-1935

Yang C, Song J, Ferguson AC, Klisch D, Simpson K, Mo R, Taylor B, Mitsuda N, Wilson ZA (2017) Transcription Factor MYB26 Is Key to Spatial Specificity in Anther Secondary Thickening Formation. Plant Physiology 175 (1):333-350. doi:10.1104/pp.17.00719

Yang C, Xu Z, Song J, Conner K, Barrena GV, Wilson ZA (2007) *Arabidopsis MYB26/MALE STERILE35* regulates secondary thickening in the endothecium and is essential for anther dehiscence. The Plant Cell 19 (2):534-548

Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, Grompe M, Koteliansky V, Sharp PA, Jacks T, Anderson DG (2014) Genome editing with Cas9 in

adult mice corrects a disease mutation and phenotype. Nature Biotechnology 32 (6):551

Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Research 40 (12):5569-5576

Zhang H, Gannon L, Jones PD, Rundle CA, Hassall KL, Gibbs DJ, Holdsworth MJ, Theodoulou FL (2018a) Genetic interactions between ABA signalling and the Arg/N-end rule pathway during *Arabidopsis* seedling establishment. Scientific reports 8 (1):15192

Zhang J, Xie M, Tuskan GA, Muchero W, Chen J-G (2018b) Recent advances in the transcriptional regulation of secondary cell wall biosynthesis in the woody plants. Frontiers in Plant Science 9

Zhong R, Demura T, Ye Z-H (2006) SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. The Plant Cell 18 (11):3158-3170

Zhong R, Lee C, Ye Z-H (2010) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. Trends in Plant Science 15 (11):625-632

Zhong R, Lee C, Zhou J, McCarthy RL, Ye Z-H (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. The Plant Cell 20 (10):2763-2782

Zhong R, Richardson EA, Ye Z-H (2007) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. Planta 225 (6):1603-1611

Zhong R, Ye Z-H (2015) The *Arabidopsis* NAC transcription factor NST2 functions together with SND1 and NST1 to regulate secondary wall biosynthesis in fibers of inflorescence stems. Plant Signaling & Behavior 10 (2):e989746

Zulawski M, Schulze G, Braginets R, Hartmann S, Schulze WX (2014) The *Arabidopsis* Kinome: phylogeny and evolutionary insights into functional diversification. BMC Genomics 15 (1):548

Appendix

Genotyping		
Name	Sequence	AGI Code
LB1	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC	
LBb1.3	ATTTTGCCGATTTCGGAAC	
35S_For	CACAATCCCACTATCCTTCGCAAGAC	
M13F	TGTAAAACGACGGCCAG	
M13R	CAGGAAACAGCTATGAC	
saf_042509_RP	TCATGAAGAATTCCATGCCTC	AT3G62440
saf1_40262_LP	GGCATTCTCAGGGAGAAAGTC	AT3G62440
SAF1_SAIL_LP	CCAGACGAGATTATCTGCCAC	AT3G62440
SAF1_SAIL_RP	CGTCACCTAATGTTCCAAAATG	AT3G62440
SAF1_exp_F2	GTGCTCGAGGAACTAACCGT	AT3G62440
SAF1_exp_F	TCTGACACGCACTCGCAAAA	AT3G62440
AT3G62440_CRISPR_R	ATTATTGGTCTCTAAACTACGTGTTGTGCTCAAGCGTcaatctcttagtcgactctaccaata	AT3G62440
NST1_g1572_F	GCAATCAACAACTGCCACGT	AT2G46770
NST2_g452_F	CCACCGGGACTAGAACCAAC	AT3G61910
AT3G58960 LP	AGCTTCTGTCTTCTTGCCCT	AT3G58960
AT3G58960_RP	GCAGATATGGCAGAGTTCGC	AT3G58960
at3g58960_amp_FP	ATGGATAGAATCAGCAGTCT	AT3G58960
58960_qRP	TCTTAAGCGCCGGAAGAAGA	AT3G58960
at3g58920_FP	ATGGATAGGATTAGCAATCT	AT3G58920
AT3G58920_RP	GCAAAATTGGCAGAAGTCGC	AT3G58920

Expression Analysis		
Name	Sequence	AGI Code
SAF1_qF_kim	CATGACTGCAGAGATGTTTTATGTCT	AT3G62440
SAF1_qR_kim	TAGGAAACGACCTCGAATTTGC	AT3G62440
MYB26_qF_Rui	TGACCCCATTGATGTTCCCAACC	At3g13890
MYB26_qR_Rui	GGTAACGCTGGCCATTGGAGA	At3g13890
NST1_qF_Rui	GCCGTAGAATTGGGATGAGA	AT2G46770
NST1_qR_Rui	GAAGCTCCTCCGACGGGACT	AT2G46770
NST2_qF_Rui	CCAAGTCAACAACTGCCACG	AT3G61910
NST2_qR_Rui	GTCGTTCTCGCGAAATCTGC	AT3G61910
IRX1_qF_kim	GGGAAAAGAAATAGGATGGATTTACG	AT4G18780
IRX1_qR_kim	GCAAAGAAGTGAATGGGTAGACAATAG	AT4G18780
58960_qRP	TCTTAAGCGCCGGAAGAAGA	AT3G58960
58960_qLP	CGTGATGCTGATCTCAACCG	AT3G58960
58920_qRP	TCCGCGTAGATCAGAGAACC	AT3G58920
58920_qLP	GAAGCGTGGTGTCTTGGATC	AT3G58920
TGA9 q L	CCTTTGACGAGAAAGCTTCCATT	AT1G08320
TGA9 q R	TGGTTCGCCATCTCTACTCA	AT1G08320
AT5G06839LP	CTCATCGGACCATGACATACC	AT5G06839
AT5G06839RP	AATTATTGTGTGGGATGTGGC	AT5G06839
PKSP_qF	AAAGAGGGTTGCAGGGTCAT	AT5G03320
PKSP_qR	CTGGAGCTGCGTATCCCATT	AT5G03320
PCRK1_321F	GGTGTCCTGCAGTGGACAAA	AT3G09830
PCRK1_752R	TATGCCCCTGCAACCCTCTT	AT3G09830

Stefan Fairburn

Amplification		
Name	Sequence	AGI Code
NST1_Pro_F	gtttgtagagttggatcagcatc	AT2G46770
NST1_CDS_NS_R	TCCACTACCATTCGACACGT	AT2G46770
NST2_Pro_F2	caacatacgtagatacctcctcg	AT3G61910
NST2_CDS_NS_R	TCCACTACCGTTCAACAAGT	AT3G61910
AT3G62440_CRISPR_F	ATATATGGTCTCGATTGCTACTGTTCTCTCTAAGCGTgttttagagctagaaatagcaagttaaaat	AT3G62440
AT3G62440_CRISPR_R	ATTATTGGTCTCTAAACTACGTGTTGTGCTCAAGCGTcaatctcttagtcgactctaccaata	AT3G62440