

Characterising the role of *KFB^T* in *Primula vulgaris* Heteromorphic Self-Incompatibility

Calum Raine

Thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

Earlham Institute

September 2019

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Floral heteromorphy in *Primula* is the phenomenon whereby one plant species produces two flower types; pin flowers have long styles with low anthers and thrum flowers have the opposite arrangement. This system promotes outcrossing and is mediated by a hemizygous *S* locus comprising five genes, one of which encodes a Kelch repeat domain protein called KFB^T . Kelch repeat domains form β propellers for protein-protein interactions. Five Kelch motifs were identified in KFB^T alongside an N-terminal F-box domain. A total of 155 Kelch proteins were found in *P. vulgaris* and conservation of KFB^T was confirmed across fifteen other *Primulaceae* species.

Spatial and temporal expression analysis of KFB^T detected consistently high transcript quantities in the gynoecium, with increasing levels in the anthers toward flower maturation. Function of the KFB^T promoter in the gynoecium was also confirmed by using a *GUS* reporter gene assay.

Plant transformation and virus induced gene silencing were carried out with a series of KFB^T overexpression and knockdown constructs out to generate irregular floral phenotypes for functional analysis. Yeast two-hybrid screens were used to identify candidate partner proteins potentially targeted for degradation by KFB^T and these inferred a possible role in the self-incompatibility system that reinforces floral heteromorphy.

Contents

Abstract	1
Acknowledgements	5
Chapter 1	6
1 Introduction	6
1.1 Floral heteromorphy	6
1.2 Kelch F-Box Proteins	19
Chapter 2	24
2 Materials and Methods	24
2.1 Plant Material	24
2.2 Seed Sterilisation	25
2.3 Media Recipes	26
2.4 Nucleic Acid Extraction	28
2.5 cDNA Synthesis	32
2.6 Nucleic Acid Quantification	32
2.7 Nucleic Acid Precipitation	33
2.8 Polymerase Chain Reactions and Product Purification	34
2.9 Comparison of <i>KFB^T</i> in Self-Compatible and Self-Incompatible Plants	39
2.10 Screening for Presence of <i>KFB^T</i> in Flowers of Four Developmental Stages	39
2.11 Gateway Cloning	41
2.12 Transformation of <i>E. coli</i>	54
2.13 Small-Scale Yeast Transformation	54
2.14 Transformation of <i>Agrobacterium</i> Strains	55
2.15 Transformation of <i>Primula vulgaris</i> Seedlings	57
2.16 Transformation of <i>Arabidopsis thaliana</i>	59
2.17 Virus Induced Gene Silencing to Knockdown <i>KFB^T</i> Transcripts in <i>Primula</i>	61
2.18 Execution of the <i>GUS</i> Reporter Gene Assay	63

2.19 Timing and Location of KFB^T Protein Expression.....	64
2.20 Yeast Two-Hybrid	69
Chapter 3.....	73
3 Bioinformatical Analysis of the KFB^T	73
3.1 Location of the Kelch Domain in KFB^T	74
3.2 Identification of Kelch Proteins Encoded by the <i>P. vulgaris</i> Genome	75
3.3 Identifying KFB^T in Other Members of the <i>Primulaceae</i> Family	83
3.4 Homology of KFB^T to <i>Arabidopsis</i> KMD Proteins	95
3.5 Identification of Conserved KFB^T Domains via Comparative Genomics	96
3.6 Transcriptional and Translational Regulation of KFB^T	99
3.7 Differential Expression of KFB^T Across Multiple Floral Morphs.....	115
3.8 Differential Expression of <i>P. vulgaris</i> Genes In Multiple Floral Morphs.....	118
3.9 Discussion.....	126
Chapter 4.....	141
4 Spatial and Temporal Expression of KFB^T	141
4.1 Efficiency and Design of KFB^T qPCR Primers	144
4.2 Temporal Expression of KFB^T in <i>P. vulgaris</i> Thrum Flower Buds	145
4.3 Differential Spatial Expression of KFB^T Between <i>P. vulgaris</i> Morphs	146
4.4 Relative Expression of KFB^T Throughout the Flower Across Development.....	149
4.5 Screening for Presence of KFB^T in Flowers of Four Developmental Stages.....	151
4.6 Identification of a Suitable KFB^T Antibody.....	153
4.7 Investigating Spatial and Temporal KFB^T Production via Western Blot.....	155
4.8 Visualisation of KFB^T Via <i>GUS</i> Reporter Gene Assay	156
4.9 Discussion.....	158
Chapter 5.....	163
5 The Function of KFB^T	163
5.1 Comparison of KFB^T in Self-Compatible and Self-Incompatible Plants.....	163

5.2 A <i>KFB^T</i> Constitutive Overexpression Vector.....	164
5.3 A Vector Containing <i>KFB^T</i> and Its Native Promoter	165
5.4 A <i>KFB^T</i> RNAi Knockdown Vector.....	166
5.5 Screening for Transgenic <i>Primula</i>	166
5.6 Screening for Transgenic <i>Arabidopsis</i>	167
5.7 Virus Induced Gene Silencing of <i>KFB^T</i>	170
5.8 Discussion.....	176
Chapter 6.....	180
The Operation of <i>KFB^T</i>	180
6.1 Preparation of <i>KFB^T</i> Activation Domain and Binding Domain Vectors	181
6.2 Screening for Specific Interaction Between <i>GLO^T</i> and <i>KFB^T</i>	182
6.3 Yeast Two-Hybrid Screen of <i>Arabidopsis</i> and <i>Primula</i> Libraries.....	185
6.4 Discussion.....	187
Conclusions	195
Appendix A	197
A1 Obtain coding sequences from a genome using a GFF annotation file.....	197
A2 Score Alignments to Calculate Similarity Across Sequences	198
A3 Generate Consensus Sequence from Multiple Nucleotide Sequences	200
A4 Count Synonymous and Non-Synonymous Substitutions	201
A5 Dot Matrix Alignment to Identify Self-Complementary Regions.....	202
A6 Identify Stem and Loop Structures in a Nucleotide Sequence	203
A7 Calculate Average Transcripts Per Million for Every Gene	204
A8 Find Closest Protein Matches in <i>Primula vulgaris</i>	205
Appendix B	206
References.....	207

Acknowledgements

I am sincerely grateful to Dr Barry Causier for offering his knowledge and experience of the various yeast two-hybrid assays used in this project. He is an excellent teacher and kindly took time in responding to communications regarding techniques and results. I would like to thank Dr Mark Smedley for supplying the Gateway cloning vectors and Dr Sadiye Hayta for her assistance during the transformation methods. I thank the UEA and John Innes Centre horticultural teams who supported this project, with particular thanks to Pamela Wells, Timothy Wells and Catherine Taylor. Roy Dunford and the laboratory support team also provided valuable service during the first half of this project.

I thank Ali Pendle for her guidance during the antibody assays. It was always a relief when Ali agreed to be part of a project; you could blindly believe that it would see completion with her help. I also thank the Earlham Institute for their accommodation, air conditioning and free coffee – all of which made the insufferable tolerable. I am indebted to Dr Mark McMullan who initially agreed to teach me bioinformatics and consequently reshaped my career path.

I thank my supervisor, Professor Phil Gilmartin, for replying to the e-mail I sent during my first week at university. I was lucky to meet a mentor who valued attitude and was prepared to take a chance on me before others would. I thank Dr Jinhong Li for his infallible patience. We have worked closely over several years, during which I have learned a lot about perseverance and what it means to take pride in your job. I respect both of their undying optimism and enthusiasm. My colleague, Dr Jonathan Cocker, has remained a sole source of permanence throughout a time of copious change. His stoicism was greatly appreciated in an environment that very much thrives on stress.

This PhD is dedicated to Jonny Walker (1980-2018) and Lee Stephenson. Jonny was a beacon of autonomous authority who inspired trust in those who perhaps could not see the light for themselves. My thesis is a token of gratitude to Lee, who first allowed me to realise that I could turn my passion for plant science into a vocation.

Chapter 1

1 Introduction

Investigations have been carried out to characterise the function, expression and mechanisms of KFB^T; one of five proteins encoded by the *S* locus that mediates floral heteromorphy in *Primula vulgaris* (Li *et al.*, 2016). A background to heteromorphy is included here and Kelch F-box proteins are introduced. An outline of the project is also provided.

1.1 Floral heteromorphy

Many of the ~350,000 angiosperm species exhibit hermaphroditism (Chapman, 2009; Paton *et al.*, 2008). This introduces the problem of self-fertilisation, which leads to inbreeding depression due to decreasing genetic diversity within a population (Charlesworth & Charlesworth, 1987). Plants have therefore developed a number of mechanisms to promote outcrossing, such as the sequential timing and positional separation of their reproductive organs (Sprengel, 1793).

Charles Darwin noted a phenomenon called floral heteromorphy that existed in *Primula* species “to favour the intercrossing of distinct individuals” (Darwin, 1862). He observed how plants from one half of a Primrose population would produce one kind of flower (called thrum) and the other half would present a second (called pin). Thrum flowers have a short style and high anthers positioned at the mouth of the corolla. Pin flowers have low anthers with long styles that present the stigma at the corolla mouth. Darwin demonstrated that inserting a proboscis into one flower morph collected pollen in such a place to correctly deposit it on the stigma when subsequently inserted into the alternate flower morph (Darwin, 1862). These opposite floral morphologies are an example of reciprocal herkogamy and in *Primula* are reinforced by a self-incompatibility system in which a plant rejects its own pollen to further ensure cross-pollination.

Style length in *Primula* is mediated by cell elongation, with thrum style cells being twice as long as those in pin (Webster & Gilmartin, 2006). Alternatively, cell sizes are equal in the lower corolla tubes (LCT) of pins and thrums; anther height is therefore controlled by cell division instead (Webster & Gilmartin, 2006). Darwin also observed differing corolla diameters between the two floral morphs and it has since been confirmed that cells in the

thrum upper corolla tube (UCT) are wider than in pin, leading to a larger thrum flower mouth (Darwin, 1862; Webster & Gilmartin, 2006). Both flower types are of equal length and, as the LCT and UCT regions are defined by point of anther attachment, it could also be said that thrum flowers have a shorter UCT and longer LCT than pins.

Thrum pollen is more opaque, spherical and larger in comparison to the oblong-shaped pin pollen grains (Darwin, 1862). There is also evidence to suggest that pin pollen may develop earlier than thrum pollen (Burrows & McCubbin, 2018). Scanning electron microscopy by Pandey and Troughton revealed thrum pollen to have rougher exine sculpturing than pin (Pandey & Troughton, 1974). Pin plants have a globular-shaped stigma with a rougher surface due to longer papillae cells (Darwin, 1862). Lewis defined the six aspects of floral heteromorphy as pollen size, anther height, male incompatibility component, female incompatibility component, style length and stigmatic papillae length (Lewis, 1949). The latter is probably a consequence of the increased cell elongation in pin styles (Webster & Gilmartin, 2006) but may also play a vital role in pollen germination (Richards, 1993).

1.1.1 Self-Incompatibility in *Primula*

Darwin observed that *Primula vulgaris* did not set seed when covered under a net and so concluded that insects were required for pollination (Darwin, 1862). In the near-absence of any visiting bees, he proposed that these flowers must instead be pollinated by moths at night. In contrast, Darwin's self-pollination tests on *P. sinensis* did successfully produce seed (Darwin, 1862). This occurred with a 24-fold increase in pins, which he assumed was caused by anthers brushing across stigmas when the post-pollination corollas fell off in this species. Thrum self-pollination could simply have occurred via gravity acting on pollen from anthers positioned above the stigma.

Darwin initially noted that intra-morph crosses are significantly less fertile than inter-morph pollination and this self-incompatibility is even stricter in thrums (Darwin, 1862). He later argued that any breakdown in the self-incompatibility system was not a malfunction but instead perfect adaptation because it retained the possibility of self-fertilisation as a final resort in the absence of insect-mediated cross-pollination (Darwin, 1877). Darwin also noted how progeny from self-fertilised *P. veris* were weaker than those from legitimate crosses (Darwin, 1868; 1876).

The stringency of self-incompatibility varies between and even within populations of the same *Primula* species (Shao *et al.*, 2019). This variance persisted when wild plants were transplanted from their separate geographies into the same glasshouse, thereby suggesting these quantitative inter-population differences in self-compatibility could have been environmentally induced pre-transplantation or even controlled epigenetically. The authors concluded that self-incompatibility is likely mediated by multiple networked genes.

Dowrick recorded that pollen tubes of *P. obconica* usually grow ~3 mm in 24 hrs during compatible interactions and rarely penetrate the stigma during incompatible self-pollination (Dowrick, 1956). The site of self-incompatible pollen tube inhibition differs between *Primula* species and also between pins and thrums of the same species (Wedderburn & Richards, 1990). Pollen germination often stalled at the stigmatic surface but some species had multiple sites of inhibition at varying distances down the style and this led Lloyd and Webb to suggest that the *Primula* incompatibility system is different from currently known mechanisms of pollen self-recognition (Lloyd & Webb, 1992). It has even been shown that *P. oreodoxa* is entirely self-fertile (Yuan *et al.*, 2019). Only *P. polyneura* and *P. pulverulenta* demonstrated absolute self-incompatibility in both floral morphs, while thrums of nine further species and pins from five more were completely self-incompatible (Wedderburn & Richards, 1990). However, these differences may be population-specific instead of species-specific (Shao *et al.*, 2019).

Self-incompatibility systems have independently evolved multiple times across different heteromorphic angiosperm lineages (East, 1940) and they fall under two classes: gametophytic and sporophytic (de Nettancourt, 1977). Efforts to describe these mechanisms in different species have so far consistently found a male and female determinant in each class that are tightly linked in the genome and therefore inherited together (McCubbin & Kao, 2000). Gametophytic self-incompatibility (GSI) is controlled by the haploid genotype of each pollen grain and sporophytic self-incompatibility (SSI) is mediated by the parent plant genotype that produced the pollen (Pandey, 1959). In this way, *Primula* self-incompatibility is said to be sporophytic (Dowrick, 1956).

1.1.2 Gametophytic Self-Incompatibility Mechanism

Normal fertilisation involves pollen adhesion to the stigmatic papillae, grain hydration and subsequent penetration of a pollen tube into the pistil, which grows through the style to the ovary. GSI is the most common incompatibility mechanism and has been identified in

over sixty plant families (Kao & McCubbin, 1996). Symptoms of this pollen rejection response in the Solanaceae family include cell wall thickening and slow growth of the pollen tube followed by its eventual eruption in the style (Ebert *et al.*, 1989). There is also a perturbed pattern of callose plugs, usually deposited at regular intervals in the pollen tube (de Nettancourt, 1977).

S-RNase glycoproteins exhibiting ribonuclease activity have been identified as the female component of GSI in the Rosacea and Scrophulariaceae families in addition to widespread Solanaceae members including *Petunia*, *Lycopersicon*, *Nicotiana* and *Solanum* (reviewed in Silva & Goring, 2001). Their expression levels correlate with acquisition of the SI response throughout flower maturation (Cornish *et al.*, 1987). These diverse ~30 kDa enzymes demonstrate marked similarity to the active site of fungal RNase T2 (Kawata *et al.*, 1988). Different allelic forms can share as little as 38 % sequence identity (Ioerger *et al.*, 1990) but there are seven to ten consistently conserved cysteine residues (Tsai *et al.*, 1992) responsible for disulphide bridges that stabilise the tertiary protein structure (Oxley & Bacic, 1996; Ishimizu *et al.*, 1996).

Antisense inhibition of such S-RNase genes is linked with self-compatible mutants (Lee *et al.*, 1994). A plant exhibiting GSI rejects pollen that possesses the same S-RNase allele as itself and this has been proven *in vivo* via gain-of-function experiments in which misexpression of an S-RNase gene from an alternative allele resulted in rejection of non-self-pollen (Murfett *et al.*, 1994). The protein backbone determines allelic specificity (Karunanandaa *et al.*, 1994) and further structural analyses found two catalytic histidine residues to be responsible for ribonuclease activity of the protein (Huang *et al.*, 1994).

Mutation of S-RNase genes in the pollen genome did not affect self-incompatibility, showing their role solely as the female component of GSI (Dodds *et al.*, 1999). The pollen component of GSI was first identified as a novel F-box protein in *Antirrhinum* (Lai *et al.*, 2002) and studies of *Petunia* have shown three divergent sets of S locus F-box (SLF) proteins are responsible for recognising non-self pistil S-RNases (Kubo *et al.*, 2010). These SLF proteins define specificity for targeted ubiquitination and degradation to disable the S-RNases and permit cross-fertilisation (Kubo *et al.*, 2016). Investigations in the Solanaceae have suggested background modifiers may be necessary to enable self-incompatibility too (Bernatzky *et al.*, 1995), as the S-allele from a self-compatible cultivar was crossed into a self-incompatible cultivar and found to be functional (Ai *et al.*, 1991).

The gatekeeper and inhibitor models have both been proposed for GSI; the former suggests a membrane-bound receptor protein uptakes allele-specific S-RNases into the pollen tube whereas the latter predicts the uptake of all S-RNases with specific inhibition occurring within the cytosol (Silva & Goring, 2001). Although the gatekeeper model has received support from McCubbin (McCubbin *et al.*, 1997), Luu showed S-RNase accumulation in pollen tubes was not allele-specific (Luu *et al.*, 2000).

Unlike in Solanaceae members, pollen rejection in *Papaver rhoeas* occurs at the stigma surface (Franklin-Tong *et al.*, 1992). It is not mediated by RNases but 15 kDa secreted proteins that are rendered dysfunctional by mutating any of four invariant cysteine residues or the aspartic acid residing in the sixth hydrophilic loop responsible for self-incompatible pollen recognition (Kakeda *et al.*, 1998). Protein phosphorylation has been detected in the GSI pathway of poppies (Rudd *et al.*, 1996; Rudd *et al.*, 1997), with inositol triphosphate potentially inducing the characteristic Ca²⁺ wave (Franklin-Tong *et al.*, 1997) that signals programmed cell death to inhibit incompatible pollen tube growth (Jordan *et al.*, 2000).

1.1.3 Sporophytic Self-Incompatibility

The SSI response in *Brassica* occurs rapidly on the stigmatic surface and prevents pollen hydration (Dickinson, 1995). The male SSI component responsible for pollen specificity was identified as an *S* locus cysteine rich (SCR) protein highly expressed in the anthers (Schopfer *et al.*, 1999). An *S* locus receptor kinase (SRK; Stein *et al.*, 1991) and *S* locus-specific glycoprotein (SLG; Nasrallah *et al.*, 1985) were the two main female SSI candidate mediators. These polymorphic proteins are believed to have co-evolved and fall under two classes; Class II is recessive to Class I and demonstrates less variance and weaker self-incompatibility (Silva & Goring, 2001). SRK was later shown to be the single definitive pistil determinant of SSI but the response appears to be enhanced by the presence of SLG (Takasaki *et al.*, 2000; Silva *et al.*, 2001).

The THL1 and THL2 thioredoxin proteins undergo inhibitive interaction with the extracellular domain of SRK (Bower *et al.*, 1996). The SCR protein is believed to diffuse through the papillae cell wall and overcome the thioredoxin inhibitors to initiate the SSI signal cascade (Cabrillac *et al.*, 2001). The downstream cellular events are still not fully understood. Promotion of ethylene signalling has been shown to breakdown SSI (Su *et al.*, 2020) and overexpression in *Arabidopsis* of an *Armadillo Repeating Containing* (ARC) gene from *Erigeron breviscapus* has been shown to induce self-incompatibility (Chen *et al.*,

2020). Calmodulin proteins, kinase-associated protein phosphatase (KAPP) and sorting nexin 1 (SNX1) have been shown to interact with SRK too (Vanoosthuysse *et al.*, 2003). Furthermore, self-compatible mutants caused by the knockout of M-locus protein kinase (MLPK) proteins in *Brassica napus* suggests they are an upstream positive regulator of SSI (Chen *et al.*, 2019).

1.1.4 Scientific History of Floral Heteromorphy

Pin flowers acquired their name due to the appearance of their stigmas like pin heads at the corolla mouths and thrums are so called because their anthers resemble the ends of weavers' threads (Darwin, 1877). It is believed that Curtis first used these two terms in the late 1700s in writing *Floral Londinensis* (cited in Gilmartin, 2015) and Hildebrand described this as heterostyly (Hildebrand, 1866). Ganders elicited surprise at the interest given to a rare breeding system that appeared in such economically unimportant plants (Ganders, 1979). However, *Primula* species have been horticulturally cultivated since the early 17th century (Li *et al.*, 2010) and currently provide an annual turnover of \$50 million for the ornamental flower industry (Richards, 2003). Florists valued Primroses by the late 1800s and the children of Darwin's village would thread the long upper corolla tube from pin flowers into necklaces (Darwin, 1877). Ganders conceded that heteromorphy offered a remarkable example of convergent evolution and lent itself well to scientific investigation, as self-incompatibility and morphology can be visually investigated without need for intensive breeding experiments (Ganders, 1979) – or modern day genotyping.

Clusius offered the earliest documented evidence for the different *Primula* morphs in 1583 but it was Darwin who first recognised their significance in nature – though Linnaeus overlooked their importance and was not concerned by such slight variations (Linnaeus, 1792). Darwin exchanged letters with Hooker in 1860 to discuss the tendency of these two forms toward adopting the male and female roles of a dioecious state. He also mentioned that Henslow, his Cambridge mentor, had noted the phenomenon too. These early observations on heterostyly from Clusius to Darwin were reviewed in Gilmartin, 2015.

Darwin identified no intermediary morphs between the pin or thrum types, which excluded the possibility of natural variation (Darwin, 1862). He also found that one plant only ever produced one kind of flower and this remained perennially persistent. Although his correspondent, Mr Wooler of Darlington, claimed to possess a spring-blooming *Primula*

that presented the alternate flower that following autumn, Darwin believed this to be a case of poor seasonal development and not a true anomaly of the model (Darwin, 1862).

1.1.5 Original Diallelic Genetic Model for Floral Heteromorphy

In 1905, Bateson and Gregory defined that floral heteromorphy in *Primula* was controlled by a single genetic unit, which they named the *S* locus (Bateson & Gregory, 1905). They proposed that thrums were heterozygous (*S/s*) for this *S* allele and pins were homozygous recessive, which accounted for the 1:1 ratio of thrums to pins observed in wild populations. While no intermediary morphs exist (Darwin, 1877), short and long homostyles do rarely occur that present both reproductive organs at the low or high positions, respectively.

Ernst undertook vast screening experiments between 1926 and 1938 to identify individuals in which heterostyly had broken down, such as self-fertile short homostyles, long homostyles and pin plants producing large thrum-type pollen. These studies led to the accepted model that floral heteromorphy was mediated by three tightly linked components at the *S* locus: *G*, *P* and *A*, which was superimposed onto Bateson and Gregory's observation (Bateson & Gregory, 1905) of a heterozygous thrum (*GPA/gpa*) and a homozygous pin (*gpa/gpa*). The dominant *G* allele – from the German word, 'griffel', meaning stylus – was thought to suppress style length, the dominant *A* allele to promote anther height and the dominant *P* allele to control larger thrum-type pollen. Lewis and Jones proposed an order of the *S* locus subunits to be *G*, *P* and then *A*, based on the reanalysis of Ernst's original data (Lewis & Jones, 1993). Webster and Gilmartin identified the first signs of heteromorphy in developing *P. vulgaris* flower buds and suggested the *G* gene acted to suppress style length before *A* acted to elevate anthers (Webster & Gilmartin, 2006).

Crosby discovered two naturally-occurring British long homostyle populations in Somerset (Crosby, 1940) and the Chiltern Hills (Crosby, 1949). Observation of their self-fertility led him to predict the consequential extinction of pin and thrum morphs due to future replacement by homostyles. He hypothesised that homostyles would almost always self-pollinate and produce greater seed quantity (confirmed by Piper *et al.*, 1984) while outcompeting thrums to fertilise pin plants and proposed this would lead to the eventual extinction of the thrum allele and render pin flowers unable to pollinate anything due to the continuing self-pollination of homostyles.

One opposing argument claimed inbreeding depression would counteract dominion of the fitter homostyles (Charlesworth & Charlesworth, 1979). Darwin indeed demonstrated that self-crossed *P. veris* exhibited less vigour after only a few generations (Darwin, 1876). However, Crosby's theory suggested this would at least result in a mixed population of pins and homostyles (Crosby, 1949) – and his ecological data appeared to confirm this process was underway. Alternatively, Bodmer proposed that homostyles rarely self-pollinate and presented thirteen years of data to show declining homostyle frequencies (Bodmer, 1958; 1960; 1984). The debate has been revisited (Curtis & Curtis, 1985) and remains ongoing at present.

Ernst believed such homostyles were derived from mutations in heterostylous plants (Ernst, 1933) while Darwin believed they arose from hybridisation events (Darwin, 1877) and Dowrick argued that recombination within the *S* locus was responsible (Dowrick, 1956). It has since been confirmed that Ernst's assumptions (Ernst, 1928b; 1936a) were correct and mutations are the cause of these homostyle variants (Li *et al.*, 2016). Charlesworth and Charlesworth explained the much rarer occurrence of short homostyles must be a symptom of their ill fitness, otherwise they would be expected to arise as frequently as long homostyles (Charlesworth & Charlesworth, 1979).

Richards proposed the inclusion of two further genes at the *S* locus to account for the male and female incompatibility components (Richards, 1997). The only allelic combination that had not been established was *GPA/GPA* – although it was reported in *P. sinensis* – and this led Kurian and Richards to propose the existence of a lethal gene that prevented this genotype (Kurian & Richards, 1997). They also restructured the *S* locus ordering to *G*, *A* then *P* before extending the model to include seven genes.

1.1.6 Progress Toward New *S* Locus Model

Modern laboratory techniques have allowed investigators to undertake differential expression analyses (McCubbin *et al.*, 2006; Huu *et al.*, 2016), mutant analysis (Webster & Gilmartin, 2003; Li *et al.*, 2008; Li *et al.*, 2010; Cocker *et al.*, 2015; Li *et al.*, 2015), genetic mapping (Li *et al.*, 2011b) and genome assembly (Nowak *et al.*, 2015; Li *et al.*, 2016; Cocker *et al.*, 2018) to elucidate the genetic mechanisms of floral heteromorphy.

Multiple *Primula* mutant phenotypes were found to be linked with the *S* locus, such as *Sepaloid*, which produces flowers containing only sepals and carpels (Li *et al.*, 2008). The

Hose-in-Hose mutant presents conversion of sepals to petals due to upregulation of the *Primula GLOBOSA* gene caused by a retrotransposon insertion within the promoter (Li *et al.*, 2010). The *Oakleaf S*-linked mutation leads to the development of lobed leaves, and further genetic mapping placed *Oakleaf* and *GLO* on opposite sides of the *S* locus (Li *et al.*, 2008; Cocker *et al.*, 2015).

Analysis of fluorescent differential display data (Li *et al.*, 2007) and random amplified polymorphic DNA led to the identification of markers now known to border the *S* locus, such as *PvSLP1*, which is tightly linked but not expressed (Manfield *et al.*, 2005). Subsequent *in situ* hybridisation assays used these markers to locate the *S* locus near the centromere of the largest *P. vulgaris* chromosome (Li *et al.*, 2011b).

McCubbin utilised suppressive hybridisation to subtract a thrum cDNA library from that of a pin flower and highlighted eleven distinct genes that, although not encoded by the *S* locus, were potentially involved in downstream pathways (McCubbin *et al.*, 2006). In 2018, Burrows and McCubbin again tackled differential expression assays to identify 229 and 311 transcripts that were respectively upregulated and downregulated in early flower buds, with 1489 and 1612 in mature buds (Burrows & McCubbin, 2018).

Nowak identified 113 genes with morph-specific expression via RNA-Seq of *P. vulgaris* and *P. veris* (Nowak *et al.*, 2015). From these candidate heterostyly genes, particular focus was given to one sequence found to be completely silenced in pin plants. This has now been identified as the *GLO^T* gene, which duplicated from *GLO* 51.7 million years ago (Li *et al.*, 2016). The Nowak publication also signified the first draft genome for *P. veris*, assembled from combined pin and thrum reads (Nowak *et al.*, 2015). However, the Gilmartin lab answered with their dual *P. vulgaris* genome assemblies of separate pin and thrum samples; a decision that proved imperative for it allowed elucidation of the *S* locus and the groundbreaking discovery of its absence from the pin genome (Li *et al.*, 2016; Cocker *et al.*, 2018).

The *Primula S* locus was identified as a hemizygous 278 kb region containing five genes, in the order of: *CCM^T*, *GLO^T*, *CYP^T*, *PUM^T* and *KFB^T* (Li *et al.*, 2016). The *GLO* duplication and positioning of the thrum-specific *GLO^T* paralogue alongside *CYP^T* at the *S* locus was also later reconfirmed by Burrows and McCubbin (Burrows & McCubbin, 2017). The *CCM^T* protein contains a conserved cysteine motif, *GLO^T* is a paralogue of *PvGLO*, *CYP^T* is a member of the cytochrome P450 family, *PUM^T* encodes a Pumilio-like RNA-binding protein

and KFB^T is a Kelch F-Box protein (Li *et al.*, 2016). The *S* locus is flanked by two *Cyclin-like F-box* genes, *CFBL* and *CFBR*. Only *CFBL* is present and shows low levels of expression; *CFBR* from the right border of the *S* locus is thrum-specific and no transcripts have been detected (Li *et al.*, 2016).

Huu compared transcriptome sequencing data between styles and corollas of pins and thrums (Huu *et al.*, 2016). From eleven candidate genes, *CYP734A50* (*CYP^T*) was confirmed via virus induced gene silencing (VIGS) to be responsible for the style length dimorphism by degrading brassinosteroid hormones in thrum styles (Huu *et al.*, 2016) to limit cell elongation (Webster & Gilmartin, 2006). The *CYP^T* *S*-locus gene also derived from duplication – like *GLO^T* (Nowak *et al.*, 2015; Li *et al.*, 2016) – and Huu detected no *CYP^T* reads in a long homostyle transcriptome dataset, providing further proof of function (Huu *et al.*, 2016). Members of two geographically separated long homostyle populations have also been found to contain mutations in *CYP^T*; a single base substitution in exon 2 and the introduction of a premature stop codon by base insertion at exon 3 (Li *et al.*, 2016) from Somerset (Crosby, 1940) and Chiltern (Crosby, 1949) populations. Furthermore, PCR designed to genotype exon three of *CYP^T* failed to amplify in the naturally occurring long homostyle species of *P. grandis*, *P. halleri* and *P. scotica* (Huu *et al.*, 2016). This points toward consistency of the model across the wider *Primula* genus.

A rare short homostyle line occurred as a progeny plant in the three-point cross to map *Oakleaf* and *Hose in Hose* (Li *et al.*, 2015) and was found to contain a 2.5 kb insertion in exon 2 of *GLO^T* (Li *et al.*, 2016). This resulted in a frameshift mutation that indicated a role for *GLO^T* in the anther height dimorphism (Li *et al.*, 2016). In the same way that Ernst used homostyles to define the genetic components of the *S* locus, homostyles have now been used to define the molecular genetic basis of heterostyly. As long (Piper *et al.*, 1984) and short homostyles are also self-compatible, this suggests single genes could mediate the male and female components of both incompatibility and dimorphism. These findings also confirm Ernst's prediction that homostyles arise from mutation (Ernst 1928b; 1936a) and not recombination (Dowrick, 1956; Lewis & Jones, 1992) or hybridisation (Darwin, 1877) – although Darwin's words preceded the rediscovery of Mendel's work and the definition of genetics.

1.1.7 Evolution of Heteromorphy in the *Primulaceae*

Heteromorphy has been observed in 199 genera across 28 angiosperm families (Ganders, 1979) over 18 orders of monocots and dicots, though it is not seen in the most primitive subclasses: Magnoliidae, Alismatiales, Hamamelidae, Arecidae and Commelinidae (Cronquist, 1968). Approximately 92 % of the ~430 *Primula* species have distylous populations and so do 40 of 41 *Dionysia* species, one of two *Hottonia* species and the single *Vitalania* species within the broader *Primulaceae* family (Richards, 2003).

The existence of heteromorphy in distantly related families suggests a polyphyletic origin with independent evolution multiple times across the plant kingdom (Lloyd and Webb, 1992), however heteromorphy only evolved once within the *Primulaceae* family itself (Mast *et al.*, 2006). The new *S* locus model (Li *et al.*, 2016) should therefore be applicable to all heteromorphic *Primula* species. The entire hemizygous *S* locus has been confirmed in *P. veris* (Cocker *et al.*, 2018).

In 1877, Darwin proposed that reciprocal herkogamy evolved before self-incompatibility as he saw no benefit to a mutation that rendered plants unreceptive to half of their population (Darwin, 1877). Charlesworth and Charlesworth used computer simulations based on the assumption of a self-compatible monomorphic ancestor (Charlesworth & Charlesworth, 1979). They suggested an initial mutation that generated pollen that was incompatible with all stigmas, before evolution drove selection for a novel stigma that was compatible with this mutant pollen. They believed this sequence of events was imperative because a novel stigma would almost certainly be eliminated without a compatible pollen type. Tight linkage would be essential for spreading the novel mutations throughout the population and the reduced rate of self-pollination would lead to the development of spatial separation between anthers and stigmas (Charlesworth & Charlesworth, 1979).

Ganders acknowledged that Charlesworth and Charlesworth's 1979 model proved mathematical viability for the evolution of floral heteromorphy but noted their ignorance toward asymmetric pollen flow caused by stamen and style dichogamy (Ganders, 1979). Ganders fundamentally disagreed with Charlesworth and Charlesworth's founding assumption of a monomorphic self-compatible ancestor due to the high rate of self-pollination and subsequent inbreeding depression that would ensue (Piper *et al.*, 1984; Ganders, 1979). Richards also contradicted Charlesworth and Charlesworth's proposal that self-incompatibility arose before dimorphism by highlighting examples of heteromorphic

plants lacking self-incompatibility yet the absence of any monomorphic plants with two mating types (Richards, 1986; 1993). Ganders proposed it was more likely a population experienced a sudden increase in self-fertilisation due to a rapid depletion of fauna (Ganders, 1979). He doubted that reciprocal herkogamy evolved to reduce self-pollination and instead proposed that the monomorphic ancestor already had spatial separation of the reproductive organs, as many plants do.

Lloyd and Webb noticed heterostyly had only evolved in tubular flowers and suggested that a self-fertile ancestor exhibiting approach herkogamy (with the stigma positioned above the anthers) acquired a dominant mutation to shorten the style, which would be selected for due to increased male fitness (Lloyd & Webb, 1992). A subsequent mutation would raise anthers to the previous stigmatic position before adapting self-incompatibility. This Lloyd and Webb model is therefore coherent with propositions from Charlesworth and Charlesworth for the development of heterostyly but instead places incompatibility at the final stage of heteromorphic evolution. Predictions by Al Wadi and Richards agree that self-incompatibility was the final aspect of heterostyly to evolve, with an initial short style mutation followed by pollen heteromorphy (Al Wadi & Richards, 1993).

However, Mast carried out phylogenetic analyses on chloroplast DNA from 207 samples (comprising 51 % of species from 95 % of the *Primula* sections) and confirmed that *Primula* heteromorphy arose from a distylous ancestor (Mast *et al.*, 2006). They showed how long homostyles first developed (as seen in *P. sinensis* and *P. verticillata*) before the thrum morph was acquired (evident in *P. boveana*, *P. gaubeana* and *P. davisii*) and high anthers were finally lost in pin morphs of *P. gaubeana* and *P. davisii*. They concluded that there are now four *Primulaceae* lineages of heteromorphy in *V. primulifora*, *H. palustris*, *P. prolifera* and the most recent common ancestor of *Primula*.

The high density of transposable elements and repetitive sequences at the *S* locus suggest it could be involved in non-homologous recombination with other parts of the chromosome (Huu *et al.*, 2016; Li *et al.*, 2016; Burrows & McCubbin, 2017). Therefore, instead of stepwise development, the *S* locus may instead have derived from segmental duplication with subsequent loss of intervening genes (Kappel *et al.*, 2017). Duplication of *CFB* genes that border the *S* locus could have initiated this illegitimate crossing over, and gene duplications at *S* loci in *Fagopyrum* and *Linum* suggest this model may apply more widely too (reviewed in Kappel *et al.*, 2017).

1.1.8 Floral Heteromorphy in Other Species

After his initial *Primula* observations in 1862 (Darwin, 1862), Darwin studied *Linum grandiflorum* (Darwin, 1863) and tristylous *Lythrum salicaria* (Darwin, 1864) before publishing his landmark text regarding the different forms of flowers on plants of the same species (Darwin, 1877). Darwin wrote how realising the purpose of heterostylous flower structure was his most pleasurable and satisfying work from a long and prestigious career (Darwin, 1887). Hildebrand first used the term 'heterostyly' and Darwin restricted its use to only those plants that contain a self-incompatibility mechanism (Hildebrand, 1866; Darwin, 1887).

Heteromorphy is presented inconsistently between species. Ganders noted that *Armeria maritima* has self-incompatibility without heteromorphy (Ganders, 1979), *Linum grandiflorum* has self-incompatibility with dimorphic styles but monomorphic anthers and *Narcissus tazetta* also presents these two features unlinked to each other (Dulberger, 1964). Flowers of *Eichhornia paniculata* present a third form (an example of tristily) that is mediated by two separate loci (Arunkumar *et al.*, 2017).

Research into *Linum grandiflorum* identified twelve morph-related genes, four of which are post-transcriptionally regulated and believed to mediate heterostyly (Ushijima *et al.*, 2012). One of these four candidates was an *S*-linked gene called *TSS1* found to be uniquely expressed in thrum styles and the authors showed that a second, *LgMYB21*, reduced pistil length when overexpressed in *Arabidopsis*. It has been suggested that the *Fagopyrum esculentum* and *Linum usatissimum* *S* loci are hemizygous too (Kappel *et al.*, 2017), as in *Primula* (Li *et al.*, 2016).

A large 5.4 Mb region containing 32 predicted genes and many transposable elements was shown to be missing from the *F. esculentum* pin genome (Yasui, 2016). Four thrum-specific transcripts were identified and a mutation in one of them led to self-compatible long homostyle plants (Yasui *et al.*, 2012), which is coherent with the behaviour of *CYP^T* in *P. vulgaris* (Li *et al.*, 2016). This candidate, *S-LOCUS EARLY FLOWERING 3 (S-ELF3)*, also arose via gene duplication from a *Fagopyrum* homologue of *EARLY FLOWERING 3 (ELF3)* in *Arabidopsis* (Yasui *et al.*, 2012).

Another well-studied heteromorphic species is *Turnera subulata*, which presents bowl-shaped flowers and not the tubular ones discussed by Lloyd and Webb in 1992. Two genes external to the *S* locus, coding for polygalacturonase (Athanasίου *et al.*, 2003) and α -

dioxygenase (Khosravi *et al.*, 2004), both demonstrated thrum-specific expression – the former is *S*-linked and the latter unlinked (Athanasίου & Shore, 1997). Another non-*S*-linked polygalacturonase gene was specifically detected in thrum styles of *F. esculentum* (Takeshima *et al.*, 2019). Labonne discovered three *S*-linked markers in *T. subulata* (Labonne *et al.*, 2009) that were used to characterise deletion mutants (Labonne *et al.*, 2010) and initiate a chromosome walk for progression toward identifying genes are the *T. subulata S* locus (Labonne & Shore, 2011).

1.2 Kelch F-Box Proteins

The purpose of this PhD project was to characterise *KFB^T* from the *P. vulgaris S* locus and identify its role in floral heteromorphy. This gene encodes a member of the Kelch F-box protein family that functions in targeted protein degradation via the ubiquitinase pathway, which is one of two major eukaryotic protein turnover systems; the other is the lysosomal pathway in which degradation is carried out by proteolytic enzymes in membrane-bound vacuoles or lysosomes (Hassan *et al.*, 2015).

1.2.1 The Ubiquitinase Pathway

Up to 50 % of plant proteins are turned over every week (Vierstra, 1993) and approximately ~5 % of the *Arabidopsis* proteome (>1300 genes) is dedicated to the ubiquitinase pathway (Vierstra, 2003). Ubiquitin is an 8.6 kDa protein discovered in 1975 (Goldstein *et al.*, 1975) that behaves as a reusable recognition tag for degradation by the 26S proteasome (Ciechanover & Iwai, 2004). This highly conserved protein is abundantly present in all studied eukaryotes and has an almost invariant 76 amino acid sequence with only rare cases of deviation outside of the plant kingdom (Catic & Ploegh, 2005). Overall stability of the structure is conferred by many hydrogen bonds that permit recycling (Vijay-Jumar *et al.*, 1987). Love found that substituting two amino acids of ubiquitin introduced destabilising cavities into the protein (Love *et al.*, 1997).

Covalent conjugation of ubiquitin to the target protein is an ATP-dependent process requiring three enzyme groups (Hassan *et al.*, 2015). An E1 ubiquitin-activating enzyme catalyses bond formation between ATP and ubiquitin before subsequently binding to ubiquitin itself (Hassan *et al.*, 2015). This activated ubiquitin is transferred to an E2 ubiquitin-conjugating enzyme and an E3 ubiquitin-protein ligase acts as the recognition

element to complete delivery of this ubiquitin complex to the target protein (Hassan *et al.*, 2015). Addition of a single ubiquitin molecule in this way is important for triggering endocytosis, virus budding or regulating gene expression via monoubiquitination of histones (Hicke, 2001). However, polyubiquitination is more common and ubiquitin is here added to one of seven lysines on a previously conjugated ubiquitin molecule (Peng *et al.*, 2003).

There are only two E1 isoforms in *Arabidopsis* and their efficiency is adequate to ensure sufficient abundance of activated ubiquitin (Pickart, 2001). Alternatively, there are 37 isoforms of E2 enzymes across 12 protein subfamilies in *Arabidopsis* (Vierstra, 1996; Backmair *et al.*, 2001). There are over 1300 E3-encoding genes in *Arabidopsis* that fall into four categories: HECT, RING, APC and SCF (Hassan *et al.*, 2015). The SCF multi-protein complex consists of SUPPRESSOR OF KINETOCHORE PROTEIN 1 (SKP1), CELL DIVISION CONTROL PROTEIN 53 (CDC53, or Cullin 1) and an F-box protein (Feldman *et al.*, 1997).

1.2.2 The F-box Domain

The F-box was initially discovered in 1996 as a 40-50 amino acid motif in humans (Bai *et al.*, 1996) that confers target specificity (Skowyra *et al.*, 1997) of the Cullin-SKP1 complex, which provides ubiquitin transferase activity. It was subsequently identified in the plant protein, UNUSUAL FLORAL ORGANS (Hepworth *et al.*, 2006), and is now known to comprise one of the largest plant protein families; there are ~700 *Arabidopsis* F-box proteins (Gagne *et al.*, 2002) and only 11, 36 and 326 in yeast, humans and nematodes, respectively (Kipreos & Pagano, 2000).

F-box proteins in plants are associated with hormone signalling, circadian clock, flowering time and pathogen defence (Kipreos & Pagano, 2000). These proteins contain F-box motifs at their N-terminus and protein-protein interaction domains at their C-terminus to confer target specificity, such as leucine rich repeats, WD domains, tetratricopeptide repeats, Armadillo domains or Kelch motifs. The SCF complex transfers target proteins to the 26S proteasome, which is a 2MDa ATP-dependent proteolytic complex consisting of 31 subunits arranged into a 20S core protease and 19S regulatory particle (Smalle & Vierstra, 2004). The regulatory particle recognises and unfolds ubiquitinated substrates before directing polypeptides into the core protease lumen (Hartmann-Petersen *et al.*, 2003), which is a cylindrical stack of four heptameric rings that cleaves peptide bonds and completes degradation (Voges *et al.*, 1999).

1.2.3 The Kelch Propeller

Kelch motifs were first discovered in *Drosophila* through mutants that resulted in female sterility due to abnormal eggs that failed to transport the nurse cell into the oocyte via the cytoplasm (Schüpbach & Wieschaus, 1991). This repeat-containing gene was later characterised and found to produce a component of ring canals that form intercellular bridges in egg chambers (Xue & Cooley, 1993). Full crystal structure analysis of the human Keap1 Kelch protein is now complete (Li *et al.*, 2004).

Kelch domains are widely dispersed throughout the biological kingdoms and initial screens found they typically contain four to seven repeats of 44-56 amino acid motifs that encode a cylinder of β -barrel sheets arranged into a propeller (Bork & Doolittle, 1994). Motifs within a protein share just 25-50% similarity (Xu & Cooley, 1993) and motif similarity between different proteins may be as low as 11 % (Bork & Doolittle, 1994). Kelch proteins have been associated with cell morphology, gene expression, cellular organisation and extracellular functions (Adams *et al.*, 2000).

While interaction sites in Kelch proteins are poorly studied, propeller proteins without Kelch domains have varying and sometimes multiple binding positions on the upper propeller surface (Lambright *et al.*, 1996), grooves between blades (Haar *et al.*, 1998; Sun *et al.*, 2017) or on opposing sides of the propeller (Renault *et al.*, 1998). It has been predicted that inter-blade loops and intra-blade variation may aid in defining target specificity (Adams *et al.*, 2000) and each propeller blade could exhibit a particular function (Schumann *et al.*, 2011).

Kelch domains paired with F-box domains are commonly but almost exclusively found in plants, with few exceptions in other organisms (Hasan *et al.*, 2015). The evolution of Kelch F-box proteins occurred prior to the divergence of animals and plants before undergoing rapid gene duplications in the plant kingdom (Sun *et al.*, 2007). They have a particularly well-characterised role in regulation of the circadian clock and flower timing. The CONSTANS flowering time regulator is stabilised via degradation of its repressors (Imaizumi *et al.*, 2005) by ZEITLUPE (ZTL; Somers *et al.*, 2000), FLAVIN BINDING KELCH REPEAT F-BOX1 (FKF1; Nelson *et al.*, 2000) and LIGHT OXYGEN OR VOLTAGE KELCH PROTEIN 2 (LKP2; Schultz *et al.*, 2001). These proteins have a C-terminal Kelch domain containing six repeats, a central F-box domain and an N-terminal LIGHT OXYGEN OR VOLTAGE (LOV) domain.

There are ~100 Kelch F-Box proteins in *Arabidopsis* (Zu *et al.*, 2009; Sun *et al.*, 2007; Schumann *et al.*, 2011) with a further three in *Juglans regia* (Yan *et al.*, 2019), 36 in *Vitis vinifera*, 44 in *Sorghum bicolor*, 46 in *Selaginella moellendorffii* and 71 in *Physcomitrella patens* (Schumann *et al.*, 2011). Between 35 and 68 Kelch F-Box proteins have been identified in *Oryza sativa* and ~35 in *Populus trichocarpa* (Schumann *et al.*, 2011; Zu *et al.*, 2009). The number of repeats is more variable in plant Kelch domains and only a single motif has been detected in some, such as the BIG24.1 protein involved in the grapevine stress response (Pasquis *et al.*, 2011). The *Chickpea F-Box 1* (*CarF-Box1*) gene also encodes a Kelch protein involved in abiotic stress (Jia *et al.*, 2012).

Of the ~35 rice Kelch F-box proteins (Schumann *et al.*, 2011; Zu *et al.*, 2009), RICE KELCH CONTAINING F-BOX 12 has been shown to modulate ethylene levels to regulate leaf senescence and grain size (Chen *et al.*, 2013), another changes panicle architecture with consequential effects on grain yield (Li *et al.*, 2011a) and IBF1 inhibits brown pigmentation in rice hull furrows (Shao *et al.*, 2012). Furthermore, hypocotyl elongation in *Arabidopsis* is controlled by the COP9 INTERACTING F-BOX KELCH 1 (Franciosi *et al.*, 2013) protein and TOO MUCH LOVE maintains root nodulation in *Lotus japonicus* (Takahara *et al.*, 2013).

1.2.4 The *KFB^T* Gene at the *Primula S* Locus

The closest homologues of the *P. vulgaris KFB^T* gene are found in the *Arabidopsis* family of four *Kiss Me Deadly* (*KMD*) genes. Expression of the *KMD* genes varies across flowers, shoots and roots but *KMD2* exhibits the highest transcript levels of all known *Arabidopsis* Kelch F-Box genes in root meristems and *KMD3* demonstrated the greatest expression in young flower tissue (Sun *et al.*, 2007). These proteins modulate cytokinin signalling via degradation of type-B *Arabidopsis* response regulators (Kim *et al.*, 2013a). Furthermore, upregulation of *KMD3* has been shown to promote root susceptibility to nematode attack (Curtis *et al.*, 2012). A Kelch F-box protein in sugar beet is also known to promote viral infection by suppressing pathogen resistance (Thiel *et al.*, 2012). An additional function for *KMD2* has been found in modulating phenylpropanoid biosynthesis by degrading the phenylalanine ammonia lyase enzyme (Zhang *et al.*, 2013). This is also known to be regulated by a second *Arabidopsis* Kelch F-box protein, SMALL AND GLOSSY LEAVES 1 (Yu *et al.*, 2019). The overexpression of *KMD1*, *KMD2* or *KMD4* is associated with a reduction of *Arabidopsis* lignin content (Zhang *et al.*, 2013).

The *KFB^T* gene was subjected to bioinformatical and experimental analyses to improve understanding of its role in floral heteromorphy. Kelch motifs were identified in the *KFB^T* amino acid sequence and comparative genomics was used to assess similarity across sequences from other *P. vulgaris* Kelch proteins, the *Arabidopsis* KMD sequences and *KFB^T* from multiple *Primulaceae* species. The *KFB^T* nucleotide sequence was screened for potential promoter elements and secondary structures that may potentially regulate transcription or translation.

Both RNA-Seq and qPCR were utilised to quantify differential *KFB^T* expression across various *P. vulgaris* floral morphs. Spatial and temporal expression of *KFB^T* was also measured across the floral whorls throughout development. These assays were supported at the protein level via transformation of a *GUS* reporter gene construct into *Arabidopsis*. Further overexpression and knockdown constructs were transformed into plants for generating mutants to aid the elucidation of *KFB^T* function. Yeast two-hybrid screens of protein libraries were used to identify partner proteins that *KFB^T* may target for degradation via the ubiquitinase pathway.

Chapter 2

2 Materials and Methods

This chapter outlines the materials and methods that were used throughout the project. Full equipment and reagents lists are provided in Appendix B.

2.1 Plant Material

Unless stated otherwise, all *Primula vulgaris* and *P. veris* plants were from glasshouses at the University of East Anglia, Norwich. They were grown under long day conditions with 16 h light periods and maintained at daytime and night time temperatures of 15 °C and 16 °C, respectively. For transformation via vacuum infiltration (Chapter 2.15.1), wildtype *P. vulgaris* seeds were purchased from Thompson & Morgan, Ipswich. With permission from Norfolk County Council, wildtype *P. vulgaris* flowers for the qPCR assays (Chapters 4.2-4.4) were selected from a roadside verge in Ketteringham, Norfolk. These were transported on dry ice to long term storage at -70 °C.

The VIGS experiment (Chapter 5.7) utilised *P. vulgaris* plants of the long-stemmed, yellow-flowered, Polyanthus horticultural variety purchased from Wymondham Garden Centre in Norfolk. The Columbia Col-0 ecotype of *Arabidopsis thaliana* were grown by Timothy Wells at the John Innes Centre and used for transformation (Chapter 2.16).

2.1.1 Definition of *P. vulgaris* Flower Developmental Stages

To monitor expression of *KFB^T* throughout flower development (Chapter 4), a consistent method for defining maturity of *P. vulgaris* buds had to be established. They were initially categorised by size but it became obvious that, due to natural variation, buds of the same length were clearly at different stages of development. This same issue of natural variation in the length of flowers at similar stages of maturity was encountered by Webster and Gilmartin (2003). They used visible physiology to recognise six stages of development – each further split into two or three substages.

Visible physiology was therefore similarly used here to define four broader stages of flower development. Stage one buds (approximately ~5 mm) were the earliest presented by the plant that could be cleanly excised from the petiole. The calyx that sheaths these young

buds was particularly variable in length. Stage two buds had begun to swell and were typically enlarged to approximately ~10 mm. Stage three buds had elongated and were beginning to present petal tissue immediately prior to opening. Stage four samples had fully opened and were the most mature flowers on the plant.

2.2 Seed Sterilisation

All seeds were sterilised prior to germination on MS media. Larger seeds, such as those of *P. vulgaris*, had to be sterilised using bleach (Chapter 2.2.1). For the smaller seeds of *A. thaliana*, sterilisation with chlorine gas was suffice (Chapter 2.2.2).

2.2.1 Sodium hypochlorite sterilisation of *Primula* seeds

Approximately fifty *P. vulgaris* seeds from the Thompson & Morgan foil sachet were put into a 2 ml Eppendorf tube and washed in 70 % ethanol. This wash step was repeated with sterile water. After adding 1 ml of 10 % sodium hypochlorite, the tube was placed on a gently rocking platform for 20 mins at room temperature. Seeds were subsequently washed three times with sterile water before 1 ml of gibberellic acid (0.4 mg/ml) was added.

After being kept in a dark place at room temperature overnight, the seeds were poured onto tissue to soak up the gibberellic acid in a sterile laminar flow hood. Sterile tweezers were subsequently used to scatter the seeds onto Murashige and Skoog (MS) media (Chapter 2.3.1). Plates were kept in darkness at 4 °C for one week before being moved to a SANYO growth cabinet with 15 h light periods, a daytime temperature of 18 °C and a night time temperature of 13 °C.

2.2.2 Chlorine gas sterilisation of *Arabidopsis* seeds

To sterilise *A. thaliana* seeds prior to germination, a glass beaker containing 100 ml of sodium hypochlorite was placed into a bell jar stored in a fume hood. The bell jar had previously been checked with the smoke from burning paper to ensure it was airtight. The seal was optimised by coating the rim of its lid in petroleum jelly. A rack of open 1.5 ml Eppendorf tubes was placed alongside it, in which the *Arabidopsis* seeds had been poured. It was essential that only a light dusting of seeds was used here to reduce crowding and ensure the maximum surface area of each seed was in contact with air. A 3 ml aliquot of

hydrochloric acid was added to the sodium hypochlorite before rapidly closing the bell jar lid. A record of the Eppendorf arrangement was kept because chlorine gas often erases ink from tubes. The bell jar remained closed overnight before the sterile seeds were scattered onto MS (Chapter 2.3.1). Plates were stored in the growth cabinet described in Chapter 2.2.1.

2.3 Media Recipes

All media was autoclaved at 121 °C for 30 mins. Any necessary antibiotics were only added after the autoclaved material had cooled below 50 °C to prevent degradation of the antibiotic compounds. Media was stored at room temperature in sterile flasks and solid media was microwaved before use, when required later.

2.3.1 Murashige and Skoog Media

MS media was used for the germination of seeds and initial growth of seedlings. Where necessary, kanamycin was added to select for transgenic plants because the vectors used for transformation contained a kanamycin resistance gene and this chemical causes necrosis of non-transformed plant tissue (Hayta *et al.*, 2018; Chapters 2.15.2 & 2.16.2). The constituents for solid MS media were as follows: 4.41 g/l of MS powder including vitamins, 30 g/l of sucrose and 8 g/l of agar. Sterile distilled water was used to achieve the desired volume and, before the addition of agar, media was brought to pH 5.8 using 3M HCl or 3M NaOH.

2.3.2 Lysogeny Broth

LB media was used in the growth of bacterial species, such as *Escherichia coli* and *Agrobacterium tumefaciens*. The constituents for LB were: 10 g/l of tryptone, 5 g/l of yeast extract and 10 g/l of NaCl. It was used both with and without 10 g/l of agar powder, depending whether the bacteria were spread on solid medium or used to inoculate a liquid culture. Antibiotics were also commonly added for selection of positive colonies.

2.3.3 Yeast Peptone Dextrose Adenine (YPDA) Media

YPDA is a general purpose non-selective growth media for yeast. It was used as an agar gel in yeast two-hybrid assays (Chapter 6), on which Y187 or AH103 yeast strains were grown

to acquire cells for transformation. YPDA liquid was used at double strength to maintain the yeast cells during the mating stage of library screening (Chapter 2.20.3). It was also used at half strength to suspend the final diploid cells for plating on selective media. YPDA was made with 20 g/l of peptone, 10 g/l of yeast extract and 20 g/l of glucose. The pH was adjusted to 6.5 before agar was added at 20 g/l (if solid plates were required) before autoclaving.

2.3.4 Synthetic Defined Medias

Synthetic defined medias were used with amino acid dropout mixes to select for plasmids in yeast two-hybrid assays (Chapter 2.20.3). The pGADT7 prey plasmid possesses a leucine biosynthesis gene (Figure 6.1 lower) and so a -leucine (-L) mixture containing all the essential amino acids except for leucine was used to select for yeast cells containing this plasmid that were autonomously providing their own leucine requirements. The pGBKT7 bait plasmid contains a tryptophan biosynthesis gene (Figure 6.1 upper) and so could be selected by using SD media made with a -tryptophan (-W) amino acid dropout mixture.

A -WL media was used to select for diploid cells containing both plasmids during the library screen (Chapter 2.20.3). This was not indicative of protein interaction but was instead used to estimate the number of screened genes and confirm successful mating. Diploid cells that did contain a pair of interacting proteins were selected with -WHL media because histidine biosynthesis was triggered by the binding proteins and both leucine and tryptophan were provided by the two vector backbones.

These medias contained 6.7 g/l of yeast nitrogen base without amino acids, 20 g/l of glucose and an appropriate quantity of amino acid dropout mixture (-W = 0.74 g/l; -L = 0.69 g/l; -WL = 0.64 g/l; -WLH = 0.62 g/l) from Clontech. The pH was adjusted to 5.8, agar was added at 20 g/l if solid media was required and the solution was subsequently autoclaved before 2.5 mM of 3-Amino-1,2,4-triazole (3-AT) was added. The 3-AT was used as an inhibitor for histidine biosynthesis and 2.5 mM was found by Dr Barry Causier (University of Leeds) to be suffice at preventing auto-activation of the *KFB^T* binding domain construct (Figure 2.3.4).

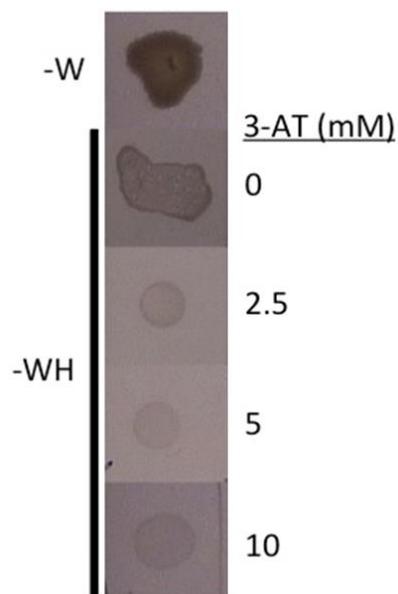


Figure 2.3.4: Auto-activation of the *KFB^T* yeast two-hybrid bait plasmid across a gradient of 3-amino-triazole inhibitor. A 2.5 mM concentration of 3-AT was found to be sufficient for preventing auto-activation of the histidine reporter gene required for survival.

2.4 Nucleic Acid Extraction

Nucleic acid extraction from plant material was used throughout the project to prepare DNA stocks for genome analysis and RNA for studying expression. A miniprep procedure was used to obtain vector constructs from bacteria and yeast to confirm plasmid integrity or identify interacting partners of *KFB^T* in mating experiments and two-hybrid library screens.

2.4.1 Plant DNA Extraction

All reagents and spin columns for DNA extraction were from the QIAGEN DNeasy Plant Mini Kit. A heating block was preheated to 65 °C. Young leaves were cut from plants, added to Eppendorf tubes and immediately frozen in liquid nitrogen before weighing to ensure they did not exceed 100 mg. A hole was first pierced into the lid of each Eppendorf to prevent tube explosion in the event that a small amount of rapidly expanding liquid nitrogen had entered the tube. Any tubes or utensils that contacted the sample during or prior to tissue disruption were also pre-chilled in liquid nitrogen. Centrifugation steps were carried out at 22000 g_n .

Frozen tissue was disrupted into fine powder using a mortar and pestle before adding 400 μ l of buffer AP1 and 4 μ l of RNase A. Samples were vortexed then incubated for 10 mins at 65 °C, during which the tubes were inverted every 3 mins. A 130 μ l aliquot of Buffer P3 was added to each sample and mixed. The tubes underwent an incubation step on ice for 5 mins.

After centrifugation for 5 mins, each sample was added to a QIAshredder spin column and centrifuged for 2 mins. The flow-through was transferred to a new tube (being careful not to disturb any pellet that may have formed) and 1.5 volumes of Buffer AW1 was added and mixed via pipetting. The samples were each passed through a DNeasy Mini spin column in 650 μ l aliquots via centrifugation steps of 2 mins, with the flow-through discarded after each spin. Collection tubes were replaced with clean ones and 500 μ l of Buffer AW2 was added to the spin columns before centrifugation for 1 min. Flow-through was discarded, another 500 μ l of Buffer AW2 was added and the samples were centrifuged for 2 mins.

Spin columns were transferred to 1.5 ml Eppendorf tubes for the final elution steps. A 100 μ l aliquot of Buffer AE was added to each spin column before incubation for 5 mins at room temperature followed by 1 min of centrifugation. This step was repeated once more. A spectrophotometer was used to measure DNA concentration (Chapter 2.6.1) and DNA precipitation was used to clean up samples and increase concentration, where necessary (Chapter 2.7). Stocks were adjusted to 100 ng/ μ l and stored at -70 °C.

2.4.2 Bacterial Miniprep

This method was used for all *E. coli* minipreps, including those for destination vectors before and after recombination during the Gateway cloning procedure (Chapters 2.11.9 & 2.11.10). It was also used to obtain crude DNA extract from *Agrobacterium tumefaciens* (Chapter 2.14.3). Centrifugation steps were carried out at 22000 g_n . All reagents and spin columns came from the Wizard Plus SV Miniprep DNA Purification System.

A 1 ml aliquot of the bacterial culture in liquid LB was used to make a 30 % glycerol stock solution for long term storage at -70 °C, when necessary. The remaining culture for each sample was centrifuged in a 2 ml Eppendorf tube before repeating this step in 2 ml aliquots until all the culture had been processed to form a pellet. The supernatant was discarded and the pellet was resuspended in 250 μ l of cell resuspension solution. Samples were not vortexed after this stage to prevent damage to the DNA.

A 250 μl aliquot of cell lysis solution was added and the tubes were inverted four times before incubation at room temperature for 3 mins. After adding 10 μl of alkaline protease, the tubes were inverted another four times before further incubation for 3 mins at room temperature. A 350 μl aliquot of neutralisation solution was added to each sample before undergoing another four tube inversions and centrifugation for 10 mins.

The liquid from each sample was transferred to filter tubes (without disturbing the pellet) and underwent centrifugation for 1 min. After discarding the flow-through, 750 μl of Column Wash Solution was added before further centrifugation for 1 min. The flow-through was discarded and 250 μl of Column Wash Solution was added to each tube. After centrifugation for 2 mins, the filters were transferred to 1.5 ml Eppendorf tubes and 35 μl of elution solution was added. Samples were left to stand for 5 mins and then centrifuged for 2 mins. This stage was repeated once. The resultant DNA concentrations were measured via spectrophotometry (Chapter 2.6.1).

2.4.3 Yeast miniprep

Yeast minipreps were used to recover vectors from yeast cells during yeast two-hybrid assays (Chapter 2.13). This allowed for PCR to be carried out on DNA extracted from the transformed yeast cell for final confirmation that the yeast two-hybrid KFB^T constructs were correct. The same technique was used later to identify cDNA fragments from potential positive yeast colonies that had displayed interaction with the KFB^T protein in library screens (Chapter 2.20.3).

Liquid yeast culture was added to a 2 ml Eppendorf tube and centrifuged for 5 mins at 22000 g_n . The supernatant was removed and these steps were repeated until the remaining sample had been processed. The final pellet was resuspended in 300 μl of extraction buffer, which had been prepared in advance of the miniprep. The extraction buffer contained 1 % SDS, 100 μM of NaCl, 10 μM of TRIS (pH 8), 1 mM of EDTA and 2 % Triton X-100.

After the addition of 0.3 g of acid washed glass beads, 300 μl of chloroform was added to the cell suspension. Samples were vortexed for 5 mins. After centrifugation for 5 mins, the aqueous phase was ethanol precipitated according to Chapter 2.7 and resuspended in a final 20 μl volume of TE buffer. This crude DNA extract was suitable for direct use in PCR

or 2 μ l could be transformed into chemically competent *E. coli* cells via heat shock for plasmid rescue (Chapter 2.12).

2.4.4 RNA Extraction

RNA extraction from *P. vulgaris* flowers was used for the spatial and temporal analysis of *KFB^T* expression (Chapters 4.1-4.5). Specimens were cut from plants and immediately transferred to pre-cooled tubes lodged in dry ice. Samples were frozen in liquid nitrogen prior to tissue disruption. All reagents and spin columns were used from the RNAqueous Total RNA Isolation Kit made by Ambion. Centrifugation steps were carried out at 22000 g_n .

Samples were weighed and these figures were used to prepare a master solution containing 12 μ l/mg of Lysis/Binding solution and 1 μ l/mg of Plant RNA Isolation Aid from Ambion. The master mix was incubated at room temperature on a roller mixer for 20 mins before use. A heating block was set at 75 °C and used to heat the final elution solution, of which 70 μ l/sample was prepared. All tubes and utensils that came into contact with the sample during or prior to tissue disruption were pre-cooled in liquid nitrogen.

Cooled pestles were fixed to a cordless pestle motor and used to disrupt each tissue sample before immediately adding the pre-calculated amount of master mix appropriate for the sample mass. Tubes were centrifuged for 3 mins and pellets were subsequently avoided with the pipette tip as each sample was added to one volume of ethanol. Samples were passed through the filter cartridge in 700 μ l aliquots and 1 min centrifugation steps.

A 700 μ l aliquot of Wash Solution #1 was centrifuged through each filter cartridge for 1 minute. This was repeated with 500 μ l of Wash Solution #1/#2. Each sample had a further 500 μ l of Wash Solution #1/#2 added before centrifugation for 2 mins. The filter cartridges were moved to 1.5 ml Eppendorf tubes and 35 μ l of pre-heated elution solution was added, allowed to stand for 3 minutes and centrifuged for 1 min. This elution step was repeated once. Nucleotide concentrations were quantified using a fluorometer (Chapter 2.6.2) and samples were precipitated to remove impurities and increase concentration, if necessary (Chapter 2.7).

2.5 cDNA Synthesis

A High-Capacity cDNA Reverse Transcription kit with random primers was used to obtain cDNA from *P. vulgaris* RNA, in coherence with the established qPCR protocol for *Primula* (Kent, 2016). Although the true quantities of gene transcripts can be misrepresented by these random primers, the amount to which they are misrepresented should be equal across samples, therefore the systematic error does not invalidate results if only one gene is being screened. This would not be suitable for comparing the expression levels of multiple genes.

An 8 µl aliquot of each RNA sample was treated with 1 µl of RQ1 DNase enzyme and 1 µl of RQ1 buffer for 30 mins at 37 °C before 1 µl of DNase stop solution was added to the mixture, which was incubated for a further 10 mins at 65 °C. An 8 µl water sample was also passed through this process, to be used as a negative control containing no template. The DNase-treated RNA was added to 10 µl of master mix containing 2 µl of buffer, 0.8 µl of dNTPs, 2 µl of random primers, 4.2 µl of H₂O and 1 µl of the reverse transcriptase enzyme. All of these components were used directly from the kit. A second negative control was setup that contained the RNA template but had 1 µl of H₂O added instead of the reverse transcriptase. This is common practice in qPCR so that the source of any contamination can be quickly localised.

A PCR thermocycler was used to incubate the samples at 25 °C for 10 mins, before raising the temperature to 37 °C for 2 hrs and finally increasing it further to 85 °C for 5 mins. The machine was chilled to 4 °C until samples could be removed. The QUBIT fluorometer was used in ssDNA mode to measure the final concentration of cDNA (Chapter 2.6.2). These were diluted for normalisation against the sample with the weakest concentration before use in PCR. The two negative control samples were not diluted.

2.6 Nucleic Acid Quantification

Nucleic acid concentrations were measured for samples after DNA extraction, RNA extraction, plasmid minipreps, cDNA synthesis and purification of PCR products. This confirmed that the processes had worked correctly, enabled standardisation of PCR template concentrations and allowed for the suitable dilutions required for DNA sequencing.

2.6.1 Nucleic Acid Quantification by NanoDrop

The NanoDrop Lite was used for spectrophotometry of DNA samples. To initialise the machine, two blank measurements were first carried out using the same buffer that the DNA sample was dissolved in, such as nuclease-free water or TE buffer. A 1.5 µl droplet was consistently used on the pedestal, which was cleaned with a lint-free tissue between measurements. The 260/280 ratio provided by the NanoDrop was used as an indication of sample purity; values of ~1.8 and ~2.0 was expected for good quality DNA and RNA, respectively (Thermo Fisher Scientific, 2012).

2.6.2 Nucleic Acid Quantification by Fluorometer

The QUBIT 2.0 fluorometer was used for the quantification of RNA after extraction or ssDNA samples after cDNA synthesis. It was also used for DNA samples when improved accuracy over the spectrophotometer was required. This increased accuracy comes from fluorescent dyes that are designed for specific binding to the desired nucleotide type being measured, thereby eliminating any sample impurities or alternative nucleotide contaminants from interfering with the quantification result.

A master mix was made containing 1 µl/sample of the appropriate fluorescent dye and 199 µl/sample of the corresponding buffer from Invitrogen®. A 199 µl aliquot of this master mix was added to each 1 µl sample to be measured. They were thoroughly mixed and allowed to reach room temperature before measurement on the fluorometer.

2.7 Nucleic Acid Precipitation

RNA, cDNA and genomic DNA samples were precipitated and resuspended in a smaller volume to increase their concentration, when necessary. This method also aids in the removal of impurities from the solution. A centrifuge was cooled to 4 °C and both 100 % and 70 % ethanol were pre-cooled in a -70 °C freezer before use. A 10 % sample volume of 3 M sodium acetate was added to each sample and thoroughly mixed before adding 3x volume of 100 % ethanol. After mixing, tubes were stored at -70 °C for 30 mins before centrifugation at 35000 g_n .

The ethanol was removed and the pellet was resuspended in 500 µl of 70 % ethanol for washing. Tubes were centrifuged for a further 10 mins (35000 g_n). The ethanol was

removed and the tubes were air dried before final resuspension to the desired volume in nuclease free water. The new concentration was checked on a spectrophotometer (Chapter 2.6.1) or fluorometer (Chapter 2.6.2).

2.8 Polymerase Chain Reactions and Product Purification

Non-quantitative PCR was used frequently for various purposes, such as acquiring amplicons for plasmid vectors (Chapters 2.11.2-2.11.6), detecting *KFB^T* transcripts (Chapter 2.10) and finding *KFB^T*-interacting partners through the yeast two-hybrid assay (Chapter 2.20.3). Quantitative PCR was used in the spatial and temporal expression analysis of *KFB^T* (Chapters 4.2-4.4).

2.8.1 Non-Quantitative PCR Reagents and Primer Design

Primers were designed manually, purchased from Eurofins Genomics and diluted to a stock concentration of 100 ng/μl for long term storage at -20 °C. They were aliquoted and diluted to a 10 ng/μl working solution before use in PCR. GoTaq Flexi DNA polymerase was used as standard. For products longer than ~1 kb – or when proofreading ability was required – Phusion High-Fidelity DNA polymerase was utilised. The dNTPs were used at a 10mM working solution.

2.8.2 Gel Electrophoresis and PCR Product Purification

To visualise PCR results, products were electrophoresed at 80 V on an agarose gel for 1 hr. The gel mixture contained 1.2 % agarose that was dissolved in 0.5x EDTA buffer under heating in a microwave before 20 μl/l of ethidium bromide was added after cooling.

Where possible, PCR product purification was carried out via the QIAquick PCR Purification Kit (QIAGEN, 2018). Centrifugation steps were carried out at 22000 g_n . The PCR product was mixed with 5 volumes of Buffer PB and centrifuged through a QIAquick spin column for 1 minute. The flow-through was discarded and 750 μl of Buffer PE was centrifuged through the spin column. After discarding the flow-through again, the column was recentrifuged for another minute to remove any excess buffer and ensure the filter was dry. The filter cartridge was transferred to a labelled 1.5 ml Eppendorf tube for final elution, in which 30 μl of Buffer EB was added to the filter disc and allowed to stand for 5

mins before the final 2 minute centrifugation step. This step could be repeated to ensure maximum product was obtained.

If multiple bands were present after electrophoresis, the desired product was excised from the gel under UV light and purified with the QIAquick Gel Extraction Kit instead. The excised gel block was weighed and 3 volumes of Buffer QG was added to 1 volume of gel. The gel was melted in the buffer via incubation at 50 °C for 10 mins with vortexing every 3 mins. After 1 gel volume of isopropanol was added to the mixture, the sample was centrifuged through a filter cartridge in 800 µl aliquots before continuing from the same Buffer PE washing step in the direct PCR purification protocol above.

2.8.3 Suitable qPCR Reference Genes

In addition to non-quantitative approaches that screened for the presence of *KFB^T* in flower buds of different sizes, qPCR was used to quantitatively compare relative expression levels throughout flower development. Template concentrations were normalised across samples to ensure fair comparison of gene expression levels. Reference genes were also selected to reinforce this, which must exhibit stable expression levels across the sample set. These provide a baseline for relative expression analysis and allow for the difference in expression quantity between a target gene and reference gene in one sample to be compared against the difference between the target and reference in another sample, which is more reliable than directly comparing raw Cq values that can be significantly affected by the error margins of pipettes and nucleotide quantification equipment. Raw Cq values are usually only used alongside standard qPCR templates to calculate the absolute nanogram quantity of target template in the sample (Life Technologies, 2011). These qPCR assays used relative expression analysis instead.

Primers had previously been designed to screen reference genes suitable for use in qPCR analysis of *P. vulgaris* (Kent, 2016). For added reliability, the *KFB^T* qPCRs underwent technical repeats with multiple reference genes: *protein phosphatase 2 (PP2A)* (F: TCATGGGTGACTATGTTGATCG; R: ATTTGCCGACTTTCGTGATTCC), *elongation factor-1α (ELF1α)* (F: TTATCGACTCGACTACTGGAGG; R: GGTAGCGTCCATCTTGTACAG) and *alpha tubulin* (F: CTATCCTTCCCCTCAGGTATCG; R: AAGCACAGCCACGTCTGTATG).

2.8.4 Efficiency and Design of *KFB^T* qPCR Primers

The SYBR Green PCR master mix was used in the *KFB^T* qPCR assays. Although this system is suitable for use in one-step qPCR – where RNA is used directly and a reverse transcription phase is added to the beginning of the thermocycling protocol – this was not used in these experiments because cDNA had already undergone synthesis and quality control previously (Chapter 2.5). The SYBR method combines both annealing and extension stages of PCR, therefore primers were designed to function at 60 °C and generate a product of ~100-150 bp from the *KFB^T* coding region. The Primer3 online primer design tool was used to highlight candidate primer pairs that fit these criteria (Koressaar & Remm, 2007; Untergasser *et al.*, 2012).

An initial non-quantitative PCR was carried out to test the suitability of six primer pairs, of which two pairs were carried forward into a qPCR trial designed to calculate primer efficiency. Primer efficiency was queried because qPCR is based on the mathematical model that template quantity doubles after every PCR cycle, therefore it had to be ensured that the primer pairs fit this model and amplified with ~100 % efficiency. This was achieved by using a trial qPCR to test the primers against a serial dilution of *P. vulgaris* cDNA. As template concentration halves, the C_q value should increase by one. Chapter 2.8.5 outlines the component quantities used in this qPCR reaction.

A primer efficiency curve was made by plotting the log of the sample dilutions against the C_q values of each primer pair. The gradient of the trendline was calculated. A slope of -3.32 indicates primer efficiency equivalent to 100 % (Thermo Fisher, 2016). The *PP2A* reference gene was used as a positive control to check that the serial dilution qPCR worked and reconfirm validity of the previously defined primers (Kent, 2016).

2.8.5 Standard qPCR Reaction

A series of qPCR assays were designed to investigate the spatial and temporal expression of *KFB^T* across various *P. vulgaris* samples, including dissected floral whorls from thrum, pin and homostyle morphs. In each investigation, sample cDNA concentration was measured on the QUBIT fluorometer (Chapter 2.6.2) and subsequently normalised to that of the weakest sample in the series. The No Reverse Transcriptase and No Template negative controls from cDNA synthesis (Chapter 2.5) were also used without dilution. Sterile distilled water was used as a third negative control.

Samples were tested for *KFB^T* expression and repeated to screen the reference gene too. They underwent three technical repeats, amounting to six wells for each sample. Individual 20 μ l reactions contained 10 μ l of SYBR Green Ready Mix, 1 μ l of the forward primer, 1 μ l of the reverse primer, 7 μ l of H₂O and 1 μ l of cDNA template. The series of qPCR assays were carried out in 96-well plates. Master mixes were made that contained all components except for the template – which was the varying component across samples – and an extra 10 % volume was added to account for pipetting error. One master mix was made with the *KFB^T* target gene primers (F: TGATTGGGACGGGATGAGT; R: CTA CTGGTGTTCGTATCCGCT) and a second was made to contain the reference gene primers.

The plate was sealed with transparent film, positioned in the qPCR thermocycler and heated to 94 °C for 2 mins before undergoing 40 cycles of 94 °C for 15 secs and 60 °C for 40 secs. After the reaction, the thermocycler carried out an automated dissociation step on the qPCR products to plot a melt curve. Monitoring peaks of this melt curve offers information about the characteristics of the qPCR products (Thermo Fisher, 2016). Multiple peaks indicate non-specific amplification and shoulders in the peak are indicative of primer dimers. This is vital to confidently rule out any spurious fluorescence in negative control samples and identify anomalous results in others. The C_q values for each sample were exported into a spreadsheet for mathematical analysis using the $\Delta\Delta$ Ct method (Livak & Schmittgen, 2001). C_q and C_t are interchangeable terms, the latter meaning ‘cycle threshold’ and will be used hereon as it is the term usually seen in mathematical publications regarding qPCR.

To calculate the expression fold change via the $\Delta\Delta$ Ct method, the individual *KFB^T* C_t value from each sample was first subtracted from its corresponding reference gene C_t value (Figure 2.8.5A). This method is regularly used to compare gene expression before and after a treatment, so at this stage the ‘before treatment’ figure is usually subtracted from the ‘after treatment’ figure. This was unsuitable for the *KFB^T* qPCR assays, so instead the figure from sample one was subtracted from each of the others, thereby presenting gene expression of each sample relative to the first (Figure 2.8.5B). These are the $\Delta\Delta$ Ct values and calculating $2^{-\Delta\Delta C_t}$ provided the final index (Figure 2.8.5C). Sample one therefore always had a value of 1, relative to itself. In this way, the reference gene provides a baseline and qPCR finds how much greater the target gene expression is over the reference gene expression in one sample, compared to how much greater that target gene expression is

over the reference gene expression in a second sample. Standard error bars were also calculated across the triplicate repeats for each sample.

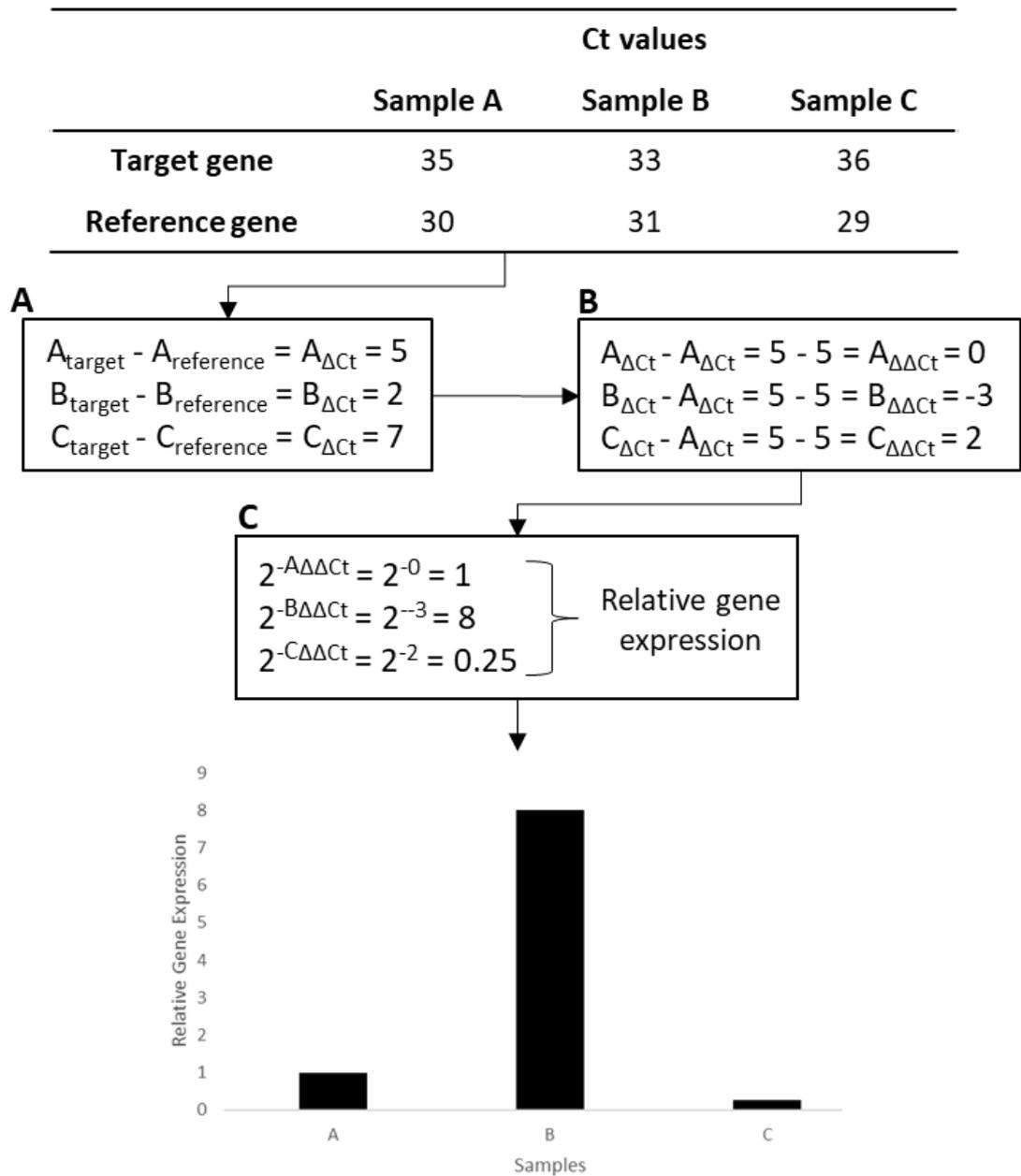


Figure 2.8.5: Outline of the $\Delta\Delta\text{Ct}$ mathematical method to calculate relative gene expression between samples via qPCR (Livak & Schmittgen, 2001). A) Reference gene Ct values are subtracted from target gene Ct values for each sample. B) The resultant value of sample A is further subtracted from each value from step (A). C) A calculation of two to the power of the negative values from step (B) provides the final relative expression index for each sample.

2.9 Comparison of *KFB^T* in Self-Compatible and Self-Incompatible Plants

To assess a possible role for *KFB^T* in the heterostylous self-incompatibility system, *KFB^T* sequences were queried in DNA samples from mutant plants identified as being self-compatible (Chapter 5.1). This screened for mutations to find whether plants that acquired a dysfunctional incompatibility system also possessed a mutated *KFB^T* sequence, which would link this gene to a role in self-incompatibility. If this was the function of *KFB^T* then mutants from the planned transgenic experiments would not present a morphologically obvious phenotype and crossing experiments would have to be carried out instead.

DNA was extracted (Chapter 2.4.1) from the young leaves of eight plants across four lines of self-compatible *P. veris* mutants. These were identified from a population that underwent mutagenesis at the International Atomic Energy Agency in Vienna for a reverse genetics experiment (Li *et al.*, unpublished). DNA was also extracted from three wildtype plants, to be used as a control in the assay. After quantification (Chapter 2.6.1), 2 µl of each sample was used in a PCR to amplify *KFB^T* from the genomic DNA. Each 40 µl reaction contained 23.6 µl of H₂O, 8 µl of Phusion High-Fidelity Buffer, 0.8 µl of dNTPs, 1.2 µl of DMSO, 0.4 µl of Phusion polymerase and 2 µl of each primer (F: ATGGAAGTTATTCTGGTCTGC; R: TCAAATTTCAACAGAACAGCCAG). These primers amplified from the *KFB^T* start codon through to its stop codon and so gel electrophoresis of the PCR products would highlight any partial or entire gene deletions.

After checking 10 µl of the PCR products on an agarose gel, the remaining 30 µl was directly purified (Chapter 2.8.2) and sequenced by Eurofins Genomics. Clustal Omega (Sievers *et al.*, 2011) was used to align *KFB^T* sequences from the eight self-compatible samples against those from the three wildtype plants to find any potential mutations between them.

2.10 Screening for Presence of *KFB^T* in Flowers of Four Developmental Stages

Assays were designed to understand the location and timing of *KFB^T* transcription, which could indicate gene function in that corresponding location and phase of heterostyly development. To understand the timing of *KFB^T* transcription, a non-quantitative PCR was carried out to initially screen for the presence or absence of *KFB^T* across flowers of four developmental stages (Chapter 4.5). This presented a broad overview of temporal *KFB^T*

expression to inform which developmental stages should be included in the downstream qPCR experiments for observing more intricate details of dynamic *KFB^T* activity.

Long homostyles were sampled in this preliminary screen due to the restricted availability of seasonal flower material at that time. RNA was extracted (Chapter 2.4.4) from each of the four developmental flowering stages of *P. vulgaris* (Chapter 2.1.1). One sample set was taken from a glasshouse long homostyle line (Chapter 2.1) deriving from a UK population in the Chilterns (Crosby, 1949) and another was taken from a Somerset population (Crosby, 1940). This process was repeated so a total of sixteen RNA samples were collected, consisting of two biological repeats of four flower stages from two long homostyle populations. There is no known phenotypic difference between the Chiltern and Somerset long homostyle lines so actually four biological repeats were made overall.

Gel electrophoresis and the QUBIT fluorometer (Chapter 2.6.2) were used to check concentration and quality of the extracted RNA before cDNA synthesis. After normalisation of the final cDNA concentrations, 1 µl of each sample was used as template in a 20 µl PCR containing 0.1 µl of Taq polymerase, 4 µl of buffer, 2 µl of MgCl₂ (25 mM), 0.4 µl of dNTPs, 11.7 µl of H₂O and 0.4 µl of each primer. The reaction was held at 95 °C for 2 mins and then underwent 35 cycles of 95 °C for 45 secs, 60 °C for 45 secs and 72 °C for 30 secs before a final extension of 72 °C for 5 mins.

Primers were used to amplify a 402 bp fragment from the middle of the *KFB^T* coding region (F: AACATTCCGAGTCGTTTCCCAAAC; R: CGTTTGGACGGCATATCTTTACC). A second PCR was simultaneously carried out on each sample using positive control primers to confirm cDNA integrity and support the *KFB^T* results (F: GTGATAATGGGACCGGAATG; R: TGCTCCGTCACAAAACAG). It was important to confirm that cDNA was the only template present, so these primers were designed across an intron in the actin gene, which proved that the earlier DNase step prior to cDNA synthesis had worked because amplification from residual genomic DNA would have given a significantly larger product. The PCR products were electrophoresed on an agarose gel and photographed under UV light to confirm the presence or absence of *KFB^T* in the sampled flower transcriptomes (Figure 4.5-2).

2.11 Gateway Cloning

Gateway Cloning was used to generate plasmid vectors for plant transformation (Figures 5.2-5.4), yeast two-hybrid screening (Figure 6.1) and Virus Induced Gene Silencing (Figure 5.7-1). These experiments were used to deduce the function and protein-protein interactions of KFB^T. Gateway cloning involves making plasmids in two stages: initial ligation of the amplified fragment of interest into an entry vector (Chapter 2.11.8) followed by recombination of the entry vector with an appropriate destination vector (Chapter 2.11.10). Entry vectors carry a Gateway cassette either side of the fragment insert that allows for recombination into the compatible Gateway cassette found on the destination vector (Figure 2.11.10). The cassettes contain *att* sites for bacteriophage lambda in *E. coli* and those on the left and right of the insert differ to preserve orientation during recombination.

A ~3 kb pCR8 plasmid (ThermoFisher: K250020) was used as the entry vector. It contains a spectinomycin resistance gene to allow for selection in *E. coli*. The kanamycin or ampicillin resistance genes are used as a selection marker in the destination vectors. The pCR8 plasmid is linearised so that only circular plasmids (formed after successful ligation with the fragment of interest) survive in *E. coli*. Destination vectors that fail to recombine do not interfere with the experiment because they contain a lethal *ccdB* gene that is otherwise swapped out for the fragment of interest during successful recombination with the entry vector. Lethality of the *ccdB* gene is derived from its interference with DNA gyrase that results in destruction of the chromosome (Bernard and Couturier, 1992).

The entry vector generation process is outlined in Figure 2.11. Once established, the entry vector acts as a checkpoint that can be kept in long term storage and used in combination with any Gateway-compatible destination vectors. Reading frame is always conserved because recombination with the destination vector is conservative in that there is no net gain or loss of base pairs (Invitrogen, 2003). All *attR* sites recombine with *attL* sites and all *attB* sites recombine only with *attP* sites. Mutations have been made in these sequences to ensure this specificity and eliminate stop codons (Invitrogen, 2003). A 43 bp fragment of the *attR* sites has also been removed to improve efficiency and ensure the reaction is irreversible. In standard gene expression, the 3'-*att* site will not be expressed and the 5'-*att* site will reside in the upstream untranslated region. Gateway cloning can also be used for gene fusion but the intervening *att* site will code for a linker of eight amino acids between the gene and its fused domain.

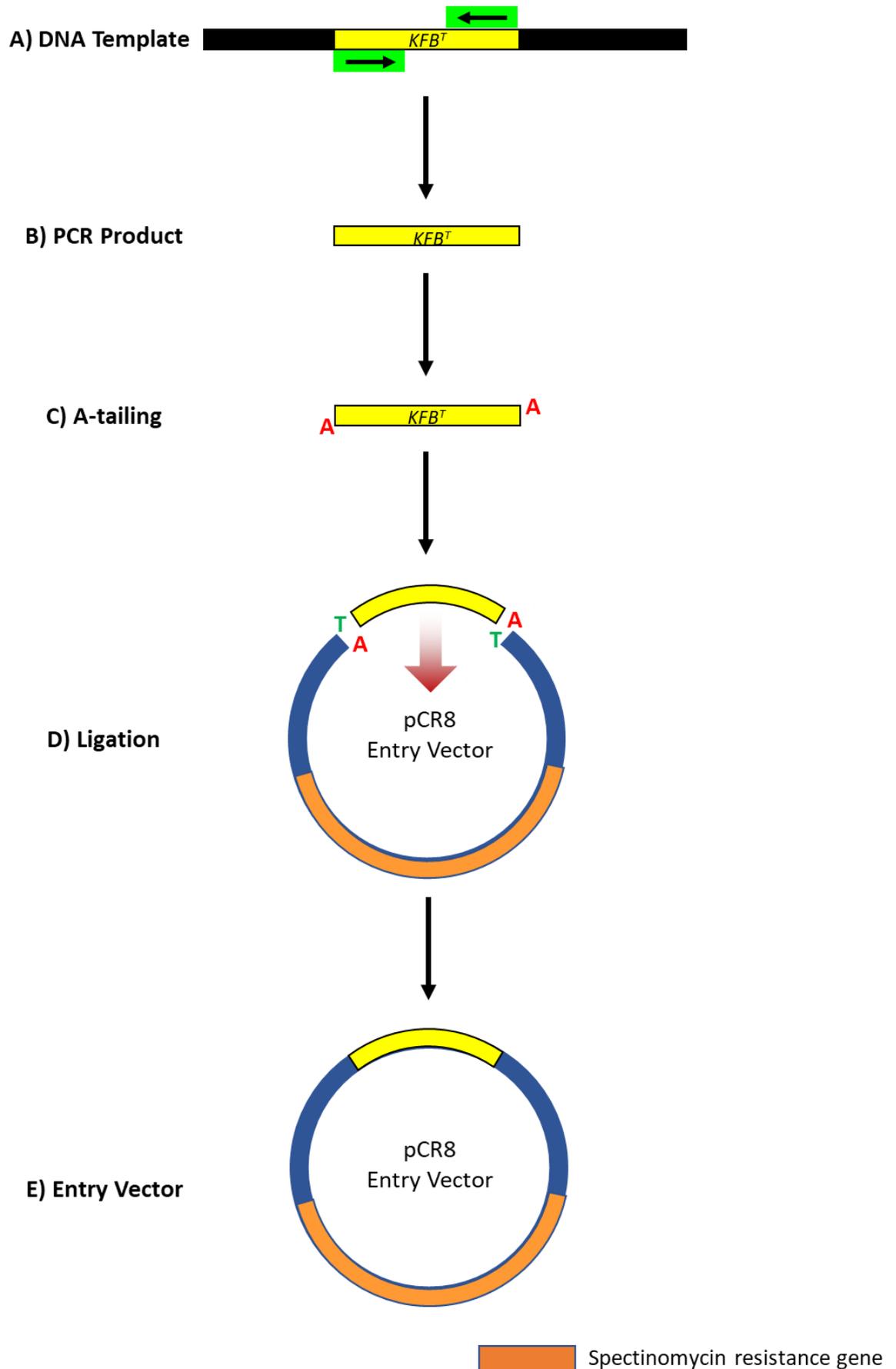


Figure 2.11: Ligation process to generate entry vectors for Gateway cloning. A) The fragment of interest is amplified. B) The PCR product is purified. C) Adenine residues are added at the 3'-ends of the purified PCR product. D) The adenine tails ligate the PCR fragment to 5'-thymine overhangs on the linearised entry plasmid. E) The completed entry vector.

2.11.1 Destination Vector Preparation

Gateway cloning can be utilised with a binary vector system in which the pSoup helper plasmid (Hellens *et al.*, 2000) carries the necessary virulence genes for *Agrobacterium*-mediated transformation and the co-transformed destination vector contains the appropriate format of genes required to generate the desired mutant. It was discovered in 1983 (Hoekema *et al.*, 1983; de Framond *et al.*, 1983) that transformation would occur successfully if this machinery was held on separate plasmids, both contained within the same *Agrobacterium* cell.

A list of destination vectors used in this project is provided in Table 2.11.1. Due to the presence of the *ccdB* gene in these plasmids prior to recombination, all destination vectors were transformed into One Shot *ccdB* Survival T1^R Competent *E. coli* Cells (Chapter 2.12). The plasmid region swapped out during the recombination reaction also contains a chloramphenicol resistance gene, so transformed cells were grown on LB agar plates containing 25 mg/ml of kanamycin and 25 mg/ml of chloramphenicol.

Table 2.11.1: Gateway-compatible destination vectors and their usage.

Destination vector	Usage	Chapter
pGWB203	<i>KFB^T</i> promoter driving <i>GUS</i> reporter gene	4.8
pBRACT114	Constitutive overexpression of <i>KFB^T</i>	5.2
pBRACT103	Transgenic expression of <i>KFB^T</i> regulated by native promoter	5.3
pBRACT507	RNAi knockdown of <i>KFB^T</i>	5.4
pTRV2	Virus Induced Gene Silencing of <i>KFB^T</i>	5.7
pGBKT7	Yeast Two-Hybrid binding domain vector (5'-fusion)	6
pGADT7	Yeast Two-Hybrid activation domain vector (5'-fusion)	6

To generate *KFB^T* fusions with the necessary transcription activation (AD) and DNA-binding domains (BD) required for yeast two-hybrid assays (Chapter 2.20), pGADT7 (AD) and pGBKT7 (BD) vectors were used. These vectors had been modified from the original Clontech backbones by Dr Barry Causier at the University of Leeds via the addition of an *att* cassette for use in the Gateway cloning system.

The pBRACT507 destination vector contains a double Gateway cassette in alternative orientations, spaced by a linker (Figure 5.4). Therefore, upon expression, the inserted fragment folds into a hairpin loop and triggers the RNA interference pathway that consequently targeted all *KFB^T* transcripts for degradation (Girin *et al.*, 2010).

The *KFB^T* coding sequence was installed into the pBRACT114 destination vector, containing the 35S promoter (Figure 5.2). The 35S promoter is from the Cauliflower Mosaic Virus (Hohn & Hohn, 1982) and is used to provide constant expression throughout an entire transgenic organism. This extreme overexpression presents the greatest chance of observing mutant effects, though it can cause non-specific or deleterious phenotypes that do not represent true native gene function (Bolognesi & Lehner, 2018).

The *KFB^T* promoter region underwent Gateway cloning into the pGWB203 destination vector, which was obtained from Dr Tsuyoshi Nakagawa at Shimane University. This vector contains a Gateway recombination site immediately preceding the *GUS* gene, directing the *KFB^T* promoter into a position in which it drives expression of the reporter. The final construct (Figure 4.8-1) therefore allowed the location of *KFB^T* activity to be visualised.

The pTRV2 destination plasmid was used for virus induced gene silencing (Figure 5.7-1). The destination vector was obtained from Professor Cathie Martin's group at the John Innes Centre and had been modified for Gateway compatibility. However, the Gateway cassette had been attached in reverse. Although this would not impede the VIGS experiment (because the fragment is reverse transcribed after expression), since the TOPO ligation occurs randomly in both directions (Chapter 2.11.8), an entry vector containing the antisense *KFB^T* amplicon was simply selected anyway – thereby being installed in the forward orientation after recombination between pTRV2 and pCR8 (Figure 2.11.10).

The viral replication machinery for pTRV2 is held on a separate pTRV1 plasmid. A stock of the pTRV1 partner plasmid transformed into the GV3101 *Agrobacterium* strain was obtained from Dr Mark Smedley at the John Innes Centre. These TRV vectors are derived from tobacco mosaic virus, which has a bipartite genome (Ramegowda *et al.*, 2014) consisting of two RNAs; one encoding the coat protein (Liu *et al.*, 2002) and the other responsible for RNA-dependent RNA polymerase, movement protein and cysteine rich protein (Macfarlane, 1999). The two vectors were only mixed at the moment of injection and this is one of the biosafety methods used to ensure containment of the virus.

Bacterial colonies containing these destination vectors were used to inoculate 5 ml of liquid LB containing 25 mg/ml of kanamycin and 25 mg/ml of chloramphenicol. They were incubated at 37 °C overnight on a platform shaking at 180 rpm. A 1ml aliquot was added to 1 ml of sterile 60 % glycerol for long term storage at -70 °C. A miniprep was carried out

on the remaining liquid culture (Chapter 2.4.2) and plasmids were digested for 1 hr at 37 °C with the *Bsr*GI restriction enzyme to check integrity of the vectors (Chapter 2.11.11).

2.11.2 Amplification of *KFB^T* Fragments for Y2H Vectors

To generate bait and prey constructs for yeast two-hybrid assays (Figure 6.1), the *KFB^T* coding region was amplified without its start codon for suitable fusion at the 5'-end with pGADT7 and pGBKT7. The *KFB^T* fragment was obtained via a 100 µl PCR containing 20 µl of Phusion High-Fidelity Buffer, 2 µl of dNTPs, 1 µl of *P. vulgaris* genomic DNA template, 3 µl of DMSO, 63 µl of H₂O, 1 µl of Phusion DNA polymerase and 5 µl of each primer (F: GAAGTTATTCTGGTCTGCCT; R: TCAAATTTCAACAGAACAGCCAG). The reaction was heated to 98 °C for 30 secs and subsequently underwent 35 cycles at 98 °C for 10 secs, 57 °C for 30 secs and 72 °C for 40 secs before a final 72 °C extension stage for 10 mins. Products were electrophoresed and excised from the gel for purification (Chapter 2.8.2).

The purified PCR product was used in the Gateway cloning procedure to generate the final *KFB^T* yeast two-hybrid vectors (Figure 6.1). Ampicillin and kanamycin were used to select for the pGADT7 and pGBKT7 vectors, respectively.

2.11.3 Amplification of *KFB^T* Coding Region for Constitutive Overexpression Vector

A vector was assembled with the pBRACT114 plasmid (Smedley & Harwood, 2014) containing the 35S promoter to constitutively drive overexpression of *KFB^T* in an effort to generate mutants and elucidate its function (Figure 5.2). The *KFB^T* coding sequence was amplified from the start codon to the stop codon for eventual recombination with the destination plasmid (Figure 5.2). To maximise amplicon concentration, a 180 µl PCR was used containing 36.13 µl of Phusion High-Fidelity buffer, 3.61 µl of dNTPs, 9.03 µl of thrum genomic DNA template, 5.42 µl of DMSO, 105.94 µl of H₂O, 1.81 µl of Phusion polymerase and 14.4 µl of each primer (F: ATGGAAGTTATTCTGGTCTGC; R: TCAAATTTCAACAGAACAGCCAG). The PCR product was directly purified (Chapter 2.8.2) and underwent Gateway cloning into pBRACT114 to generate the final overexpression vector (Figure 5.2).

2.11.4 Amplification of *KFB^T* Fragment for RNAi Knockdown and VIGS Vectors

A 549 bp fragment was amplified for use in pBRACT507 (Figure 5.4; Smedley & Harwood, 2014) and pTRV2 (Figure 5.7-1) plasmids for the elucidation of gene function via RNAi and

viral knockdown of *KFB^T*. The targeted region commenced 450 bp upstream of the *KFB^T* stop codon and continued a further 99 bp into the 3'-untranslated region (Figure 2.11.4).

```

ATGGAAGTTATTCCCTGGTCTGCCTGAAGATCTTGGACTCGAGTGTATGATTCGTTCCACAT 60
TACACAACATTCCGAGTCGTTTCCCAAACATGCCATTTATGGAGAAAACCTTCCAAACC 120
ACAGATTTTTCATAGTTACAGAAAAGAACAAGGATACAGCCACAAAATGATTTGTTTTGTA 180
CAATCTATTCCGCCAAATGCTCTTGCTGATGAAACTGGAAAATCAGCCAACCTCGTGCGGT 240
TATGGGATCACTGTATTTCGACGAAGTCGGACATGGGGGAGACTTAGTCAGGTTCCCAAGT 300
ACCAAAGCGGGCTCCCCTTGTTTTGCCGTGTGGCGAGCTCCGGAATAAGCTCATCGTTA 360
TGGGCGGGTGGGACCCGTTTAGTTACCACCCAGTTAAGGACGTGTTTGTATTGATTTTG 420
TGAACCAGTTGTGGCGACAAGGTAAAGATATGCCGTCCAAACGATCATTTTTTCGCTATGG 480
GCGCCATAGATGTTAAGGCATGTATACGTCGCGGGGAGGCCACGATGAGAATAAAGGTGCT 540
TTGAAATCAGCTTGGGTTTATGATTTGGGACGGGATGAGTGGACTGAGATGATACAGATG 600
GCACACGAACGTGACGAGTGTGAGGGAATAGTGATGGGTAATGAGTTTTGGGTAGTTAGC 660
GGATACGACACCAGTAGTCAAGGGGTGTTTGTGACGAGCGCAGAGTCCTATTGTGTCAGT 720
ACCAGGATGTGGAATCTAGTTGAATCTGTATGGAAGGCGGGCCAGTGCCCGAGGTCATCT 780
GTCCTAAGTTTTAAAACCTAGTCAGTTGATAAGTTATAACCGAATTCAGCTCGGCTATTACA 840
GATGGAGCATTTGGGATCGCGCTGGGCGGCCAGATTCTCTTAAAGGAGTCAGCAGATGTT 900
GATGTAAGAAAGCTTTTTTTTTTAGTGGATGTTGGTGAAGGGCAAACCTATAGAATCGAG 960
AAGATTAATGTGCCTGATCAGTTTTCTGGTTTAGTTCAATCTGGCTGTTCTGTTGAAATT 1020
TGATAGTTAAACGCCTTTGCTAACATTTTAAAAATGTAGTATGTGTTCTATTTGATTTTGT 1080
TTTTAAAAGAATACAAGTATAAGTAAGTACGGTTTCTACAAATAAAGAAGAAGATAGTCG 1140
TGATATTGATAAATAACATATAAATCCTTTTGGACCACAACCAACAGCCAAATAACTAAT 1200
ATGGCGATATGACCAACTTTATACGAAAACCTAGCATAGTGCCCGTGCAATCGCCCGGGT 1260
CAGCATCATAGTGTCCAATCAATGTTGCCACGATGTCATTTCTATATAAATTTTACAAT 1320
TGTAGCAATCCATCCTCTTAGCTCCTTTTCTCGCATCGT 1380
Start Codon Stop Codon Forward Primer Reverse Primer

```

Figure 2.11.4: The fragment amplified from the 3'-end of *KFB^T* for use in its knockdown via RNA interference and viral induced gene silencing.

The fragment was obtained via a 60 µl PCR containing 36.3 µl of H₂O, 12 µl of Taq Buffer, 1.2 µl of dNTPs, 6 µl of MgCl₂ (25 mM), 1 µl of *P. vulgaris* thrum genomic DNA template, 0.5 µl of Taq polymerase and 1.5 µl of each primer (F: CGGGATGAGTGGACTGAGAT; R: GTAGAAACCGTACTTACTTATAC). The reaction was heated to 95 °C for 2 mins before undergoing 35 cycles of 95 °C for 30 secs followed by 51 °C for 30 secs and 72 °C for 1 min. A final 72 °C extension step was carried out for 10 mins and samples were maintained at 4 °C until 15 µl was electrophoresed on an agarose gel. The remaining 45 µl was directly purified (Chapter 2.8.2) and used in the Gateway cloning procedure.

2.11.5 Amplification of *KFB^T* and Its Native Promoter

A Ti plasmid vector was made with *KFB^T* under regulation by its native *P. vulgaris* promoter (Figure 5.3) to observe the effects of expressing this gene at normal levels in isolation from any other *S* locus genes. This could be achieved via transformation into pin plants or *Arabidopsis*. An amplicon beginning 3 kb upstream from the start codon was used for this

vector; usually ~2 kb is suffice however initial sequence analysis suggested *KFB^T* may have a 5'-UTR in excess of 1 kb (Chapter 3.6). The fragment extended through the open reading frame and continued until ~450 bp downstream of the *KFB^T* stop codon to ensure inclusion of the transcription terminator. A total length of ~4.5 kb was therefore amplified in a 50 µl PCR containing 1.25 µl of thrum genomic DNA template, 29.85 µl of H₂O, 10 µl of Phusion High-Fidelity buffer, 1.25 µl of dNTPs, 1.4 µl MgCl₂ (50 mM), 1.25 µl of Phusion polymerase and 2.5 µl of each primer (F: GTTAATTATGGTGTGTTACC; R: CTGAAACTTCAACAACTACC). The entire PCR product was electrophoresed on an agarose gel (Chapter 2.8.2). This presented two bands, of which the upper ~4.5 kb fragment was excised from the gel and purified for use in pBRACT103 (Chapter 2.8.2).

2.11.6 Amplification of the *KFB^T* Promoter for Use in the *GUS* Reporter Gene Construct

The *KFB^T* promoter had to be amplified (Figure 2.11.6) from the *P. vulgaris* genome for inclusion in the *GUS* reporter gene vector (Figure 4.8-1) to visualise location of *KFB^T* activity. This included the 5'-UTR after the transcription start site, so the *GUS* reporter gene may be subjected to any regulatory activity that this region might natively impart on *KFB^T*. Simply amplifying 2 kb upstream of the start codon is usually suffice for obtaining a gene's promoter (Rockman & Wray, 2002), however the potentially long 5'-untranslated region of *KFB^T* (Figures 3.6.1-1 & 3.6.3-3) meant it was decided to amplify a ~3 kb region upstream of the *KFB^T* start codon instead (Figure 2.11.6).

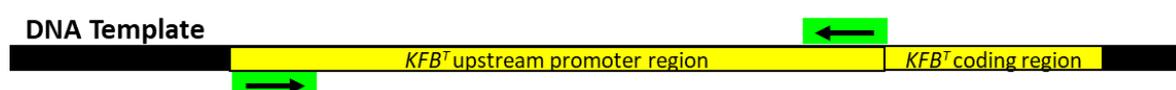


Figure 2.11.6: The 2.9 kb region immediately upstream of the *KFB^T* coding sequence was amplified for use in the promoter-driven *GUS* reporter gene construct to highlight the location of *KFB^T* protein production *in vivo*.

To improve PCR efficiency and reduce any risk of non-specific amplification from the genome, a vector containing the same desired *KFB^T* promoter region (Chapter 2.11.5) was used as PCR template. Phusion polymerase was utilised due to the requirements for high-fidelity and large product size. A 60 µl reaction volume was used, containing 34.4 µl of H₂O, 12 µl Phusion high-fidelity buffer, 1.2 µl of dNTPs, 0.6 µl of polymerase enzyme, 1.8 µl of DMSO, 3 µl of MgCl₂ (50 mM), 2 µl of DNA template (10 ng/µl) and 3 µl of each primer (F: GTTAATTATGGTGTGTTACC; R: ATTATTTTCAAAGATTGGCTATGAAATAG). The reaction was held at 98

°C for 30 secs before undergoing 45 cycles of 98 °C for 10 secs followed by 55 °C for 30 secs and 72 °C for 5 mins.

After a final extension step at 72 °C for 5 mins, products were maintained at 4 °C until 10 µl could be analysed via gel electrophoresis. Although minor non-specific bands appeared faintly on the gel, the remaining 50 µl were directly purified (Chapter 2.8.2) because the brightness ratios between gel bands showed the majority of amplicons in the product pool would be correct and proper downstream screening checkpoints were in place to confirm this.

2.11.7 A-tailing Blunt-End PCR Products

The *Taq* polymerase enzyme automatically adds an adenine residue to the 3'-end of all PCR products (Figure 2.11C) but Phusion polymerase does not. This adenine sticky end facilitates ligation with the overhanging 5'-thymine residues on the pCR8 entry vector (Figure 2.11D). The blunt-end PCR products generated from Phusion polymerase must first undergo A-tailing before their use in Gateway cloning. To achieve this, 4.4 µl of fresh PCR product was added to 2 µl of GoTaq Flexi Buffer, 1 µl of GoTaq Flexi DNA polymerase, 2 µl of dATPs (10 µM) and 0.6 µl of MgCl₂ (25 mM). The solution was incubated at 72 °C for 6 hrs.

2.11.8 PCR Product Ligation into the pCR8 Entry Vector

Cloning of PCR products into the entry vector (Figure 2.11D) utilised the pCR8/GW/TOPO TA Cloning Kit. This reaction included 1 µl of pCR8 vector and 1 µl of salt solution (1.2 M NaCl and 0.06 M MgCl₂). A volume of PCR product was calculated to allow for a 3:1 molecular ratio of insert to entry vector (Equation 2.11.8). The product concentration in Equation 2.11.8 must consider the 2.27x dilution that the PCR underwent during the A-tailing reaction (Chapter 2.11.7). The pCR8 plasmid is 2817 bp in length and Invitrogen provide this at a concentration between 5 and 10 ng/µl so a figure of 7.5 ng/µl was assumed and used in the equation. After the addition of this pre-calculated volume to the ligation mixture, the solution was diluted with water to a final volume of 6 µl. Ligation was carried out overnight at room temperature and the products were subsequently transformed into *E. coli* (Chapter 2.12).

$$\text{volume of PCR product} = \frac{(3 \times \text{pCR8 concentration}) \div \frac{\text{PCR product length}}{\text{pCR8 length}}}{\text{PCR product concentration}}$$

Equation 2.11.8: Volume of PCR product required in ligation reaction to ensure a 3:1 molecular ratio between the PCR product and the pCR8 entry vector.

2.11.9 DNA Sequencing

Fragments undergoing TA cloning can be inserted in either orientation, so the direction and integrity of these inserts had to be checked. After miniprep from *E. coli*, DNA samples were therefore sent for sequencing. Products less than 1 kb in length were provided at a concentration of 5 ng/μl and longer products were provided at a concentration of 10 ng/μl. Entry vectors were sequenced by Eurofins Genomics using primers with the standard M13 Uni (-21) sequence: TGTAACGACGGCCAGT. Vectors with larger inserts were also sequenced using primers with the standard M13 rev (-49) sequence: GAGCGGATAACAATTTCACACAGG. Plasmid samples were provided as 15 μl aliquots containing DNA at a concentration of 50-100 ng/μl. When non-vector PCR products were sequenced that did not contain a standard primer site, 2 μl of an appropriate user-defined primer (10 ng/μl) was added instead.

A chromatogram was used to allow analysis of the presence and fluorescence of each base throughout the sequencing run. This aided in identifying regions of low intensity or areas containing multiple sequences. Sequencing results were provided as raw reads and also as clipped sequences that had 5' and 3'-regions of poor quality removed. These were used in BLASTn searches to check their identity. Clustal Omega (Sievers *et al.*, 2011) online alignment software was used to screen for errors. If anomalous bases were found, it had to be established whether these were sequencing errors or if they were integral to the amplified sequence – in which case the cloning was repeated.

2.11.10 Entry Vector Recombination with the Destination Vector

To complete the Gateway cloning procedure and assemble the desired construct containing a DNA insert of interest, the entry vector was recombined with the destination vector (Figure 2.11.10). Recombination occurred between the *attL* sites of the entry vector and the *attR* sites of the destination vector in a 10 μl reaction consisting of 6 μl of TE buffer, 1 μl of destination vector, 1 μl of entry vector and 2 μl of LR clonase II enzyme. A vector (150 ng/μl) to entry vector (50 ng/μl) ratio of 3:1 was used for most recombination

reactions. A 1:1 ratio was found to be more effective in assembling the VIGS construct (Figure 5.7-1).

The recombination solution was gently mixed and incubated at 25 °C overnight before 1 µl of Proteinase K solution was added and the reaction was incubated for 10 mins at 37 °C. A 2 µl aliquot of the reaction was transformed into Subcloning Efficiency DH5α chemically competent cells, according to the heat shock method outlined in Chapter 2.12. After overnight incubation on LB agar plates containing 50 µg/ml of kanamycin (or ampicillin for the pGADT7 yeast two-hybrid vector), positive colonies were selected and used to inoculate 5 ml of liquid LB. The culture was subsequently incubated at 37 °C overnight with shaking at 180 rpm before miniprep (Chapter 2.4.2) the following day.

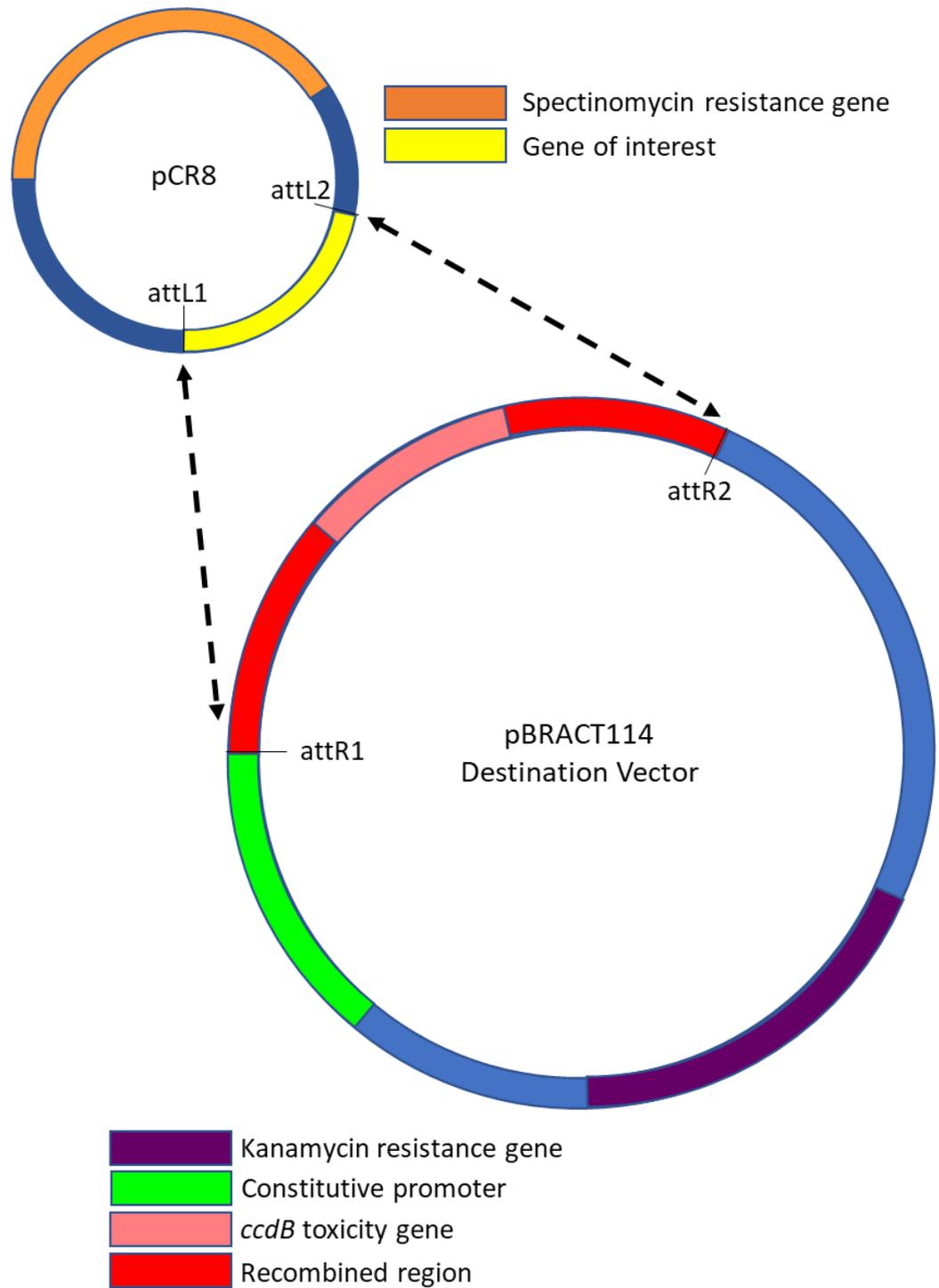


Figure 2.11.10: Recombination between the *att* sites of the entry and destination vectors during Gateway cloning. A constitutive overexpression vector is shown but the principles apply to the entire series of Gateway-compatible vectors.

2.11.11 Restriction Enzyme Digest of Plasmid DNA

After the Gateway cloning procedure, a *BsrGI* restriction digest was used to confirm integrity of the final assembled constructs (Figure 2.11.11). The expected band patterns were derived from the known vector maps and insert sequences. These predicted band patterns, plasmid concentration and total size were used to calculate the specific amount of template required for digestion to ensure the smallest band would contain 50 ng of DNA, which is the lowest amount generally visible on a gel after accounting for losses through inefficiency and error (Equation 2.11.11). The digest reactions consisted of this calculated sample quantity with 1 μ l of *BsrGI*-HF (high fidelity) restriction enzyme, 5 μ l of 10x CutSmart Buffer and H₂O to dilute the mixture up to 50 μ l. The reaction was incubated for 1 hr at 37 °C before undergoing DNA precipitation (Chapter 2.7) to clean up buffer salts and enhance the gel electrophoresis quality.

$$\text{Sample volume to digest} = \frac{50}{\text{Smallest fragment length} \div \text{Total sample length}} \div \text{Sample concentration}$$

Equation 2.11.11: Volume of sample required in digestion reaction to ensure smallest fragment is visible on agarose gel.

Digested DNA fragments were separated by size via electrophoresis (Chapter 2.8.2) and the agarose gel was subsequently photographed under UV light. The band pattern was compared against the expected pattern to confirm successful recombination of the correct fragment into the destination vector. This assay also ensured no parts of the vector backbone were lost during cloning.

The Gateway cassette contains a *BsrGI* restriction site on either side of the insert, therefore this enzyme excises the insert fragment and allows plasmid integrity to be checked by band size. However, the sequence of some Gateway destination vectors contains a single *EcoRI* restriction site in the vector backbone. The *KFB^T* sequence contains an *EcoRI* restriction site too. Therefore, using *EcoRI* resulted in asymmetric digestion of the insert fragment and allowed for the final orientation of the insert to be screened as well as just the fragment sizes. This should not have been necessary due to the specificity of the *att* recombination sites but it did implement a secondary check for earlier user error. This was useful when using pBRACT507 to generate an RNAi vector to knockdown *KFB^T* (Chapter 5.4) as it was imperative that this vector contained two copies of the insert in inverted orientations to form a hairpin loop.

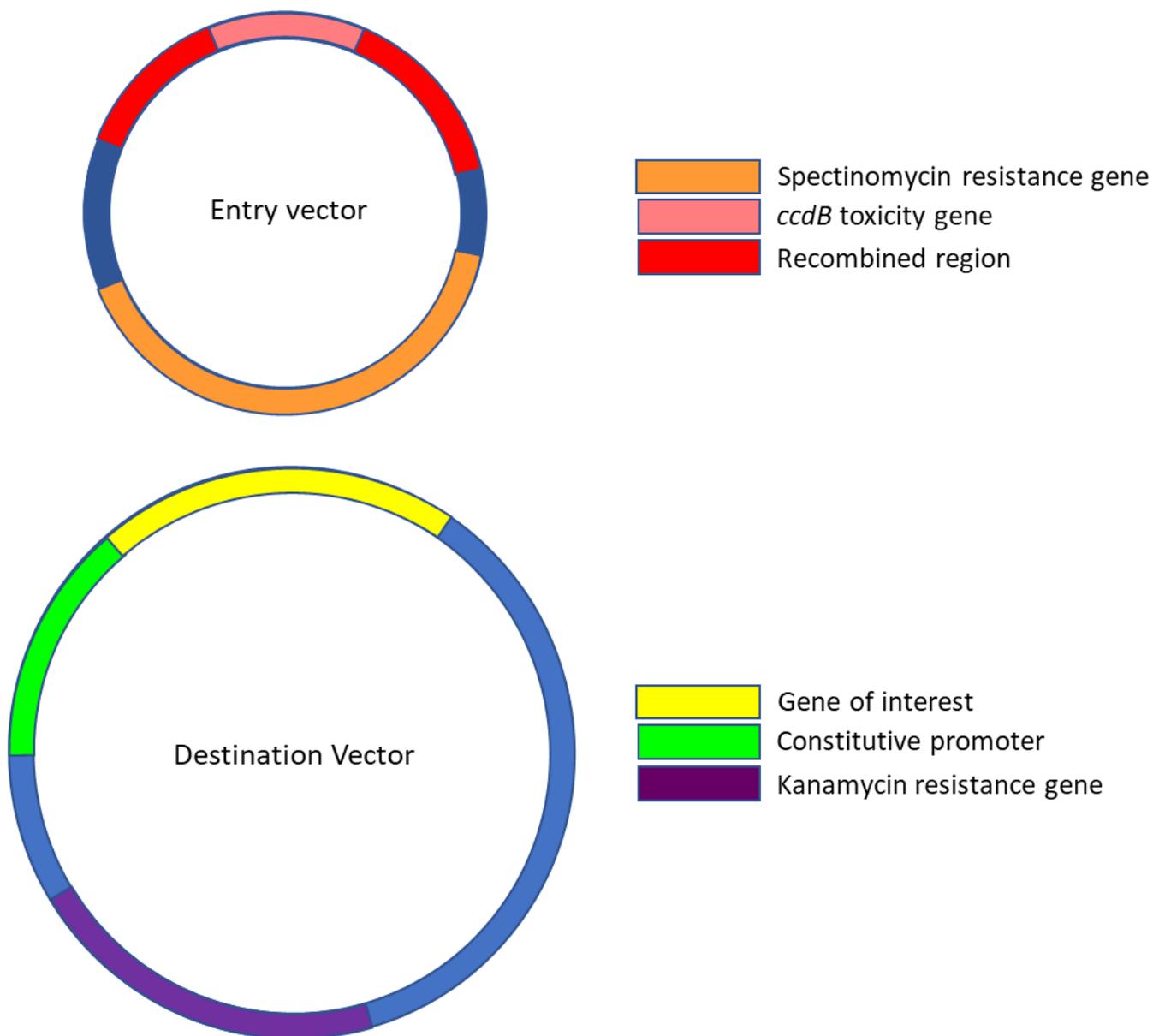


Figure 2.11.11: Products of the recombination reaction in Gateway cloning. The gene of interest is appropriately inserted into the destination vector at the 3'-end of the promoter element. A constitutive overexpression vector is shown but there are a multitude of Gateway-compatible vectors for various other functions. The recombined entry vector is discarded and the completed destination vector is transformed.

2.12 Transformation of *E. coli*

Heat shock was used to transform vectors into *E. coli* cells. One Shot TOP10 chemically competent cells from Invitrogen™ were used for entry vector transformation after the ligation reaction. For preparation of destination vectors prior to the Gateway recombination reaction, One Shot *ccdB* Survival T1^R competent cells were utilised for transformation. For all other *E. coli* transformations, including successfully cloned destination vectors, Subcloning Efficiency DH5 α chemically competent cells were used.

To transform products of the TOPO cloning reaction (Chapter 2.11.8), a water bath was heated to 42 °C and ice was used to pre-chill 2 ml Eppendorf tubes containing 2 μ l of ligated entry vector ligated with the A-tailed PCR product. After slow thawing, 50 μ l of *E. coli* cells were added to the Eppendorf containing the ligated entry vector before being briefly vortexed to gently mix. These cells were left for 30 mins on ice before being heat shocked in the 42 °C water bath for 30 secs and immediately placed back on ice.

After remaining on ice for 2 mins, 200 μ l of S.O.C. medium was added and the tube was shook horizontally at 200 rpm for 90 mins in a 37 °C incubator to allow for the cells to recover. Both 50 μ l and 150 μ l aliquots were spread onto separate LB agar plates containing 100 μ g/ml of spectinomycin. The remaining 50 μ l of transformed cells were stored at 4 °C in a refrigerator. The plates were incubated at 37 °C overnight in a static incubator. Positive colonies were used to inoculate 5 ml of liquid LB containing 100 μ g/ml of spectinomycin before further overnight incubation at 37 °C with 180 rpm shaking.

2.13 Small-Scale Yeast Transformation

The *KFB^T* bait and prey constructs (Chapter 2.11.2) for yeast two-hybrid (Chapter 6) were transformed into yeast strains Y187 and AH103, respectively. Stocks of both yeast strains were prepared for transformation by growing on YPDA media (Chapter 2.3.3) at 30 °C for three nights. Colonies from both strains become pink over time but AH109 gains colour faster than Y187, especially on media containing a low adenine quantity. This was used as visual confirmation that the yeast strains were correct before transformation.

Centrifugation steps were carried out at 22000 g_n . Salmon sperm DNA at a concentration of 2 mg/ml was boiled for 10 mins and immediately chilled on ice prior to use. A pipette tip was swiped once across the yeast culture plates to collect a cluster of cells, which were

resuspended in 1 ml of sterile distilled water. They were centrifuged for 15 secs and the pellet was resuspended in 900 μ l of H₂O and 100 μ l of 10x lithium acetate (1 M).

The solution was centrifuged again and the supernatant was removed before the following reagents were added in this order without mixing: 240 μ l of 50 % PEG, 36 μ l of 10x lithium acetate, 50 μ l of denatured salmon sperm DNA, 5 μ l of plasmid DNA to be transformed and 29 μ l of H₂O. The mixture was vortexed to resuspend the pellet and then incubated at 42 °C for 30 mins, with tube inversions every 10 mins. The purpose of the PEG was to sink to the bottom of the Eppendorf and form a barrier to limit contact between the cells and LiAc until transformation commenced after vortexing.

The transformed cells were centrifuged for 15 secs before finally being resuspended in 100 μ l of H₂O and spread on a plate for incubation at 30 °C. To select for the correct cells, pGADT7 transformants were grown on -leucine (-L) dropout media and pGBKT7 transformants were grown on -tryptophan (-W) dropout media (Chapter 2.3.4).

2.14 Transformation of *Agrobacterium* Strains

Transgenic plants were generated to deduce *KFB^T* function by monitoring the physiological effects of modified gene activity (Chapter 5). These *Primula* and *Arabidopsis* lines required the use of *Agrobacterium tumefaciens* that had been co-transformed with both pSoup and the desired destination vector. Multiple vectors had been constructed (Chapter 2.11) to analyse the function and location of *KFB^T* via the *Agrobacterium*-mediated transformation of *Primula* and *Arabidopsis*. Electroporation was used to transform these vectors into their respective *Agrobacterium* hosts.

2.14.1 Preparing Electrocompetent *Agrobacterium* Strains

The AGL1 strain was obtained from Dr Mark Smedley at the John Innes Centre in Norwich and used for the overexpression, knockdown and reporter gene constructs (Chapters 4.8 & 5.2-5.4). The GV3101 strain containing the pMP90 helper plasmid was obtained from Dr Laurence Tomlinson at The Sainsbury Laboratory in Norwich and used in the VIGS assay (Chapter 5.7).

To prepare an *Agrobacterium* stock for long term storage, one colony of the AGL1 strain was selected and used to inoculate 5 ml of liquid LB media. This was incubated for 6 hrs at

30 °C with shaking at 220 rpm before being directly poured into 100 ml of LB containing 25 µg/ml of rifampicin and 25 µg/ml of carbenicillin, which are the AGL1 selection markers.

After overnight incubation at 30 °C with shaking at 220 rpm, the culture was spun for 5 mins at 8400 g_n . The pellet was resuspended in 50 ml of ice-cold sterile 10 % glycerol. These centrifuge and wash steps were repeated three times. Cells were resuspended in 1 ml of 10 % glycerol and aliquoted into 40 µl stocks for storage at -70 °C. The same process was followed for preparation of the GV3101 *Agrobacterium* cells by Dr Jinhong Li, except gentamycin was used instead of carbenicillin to select for the Ti plasmid.

2.14.2 Electroporation of *Agrobacterium* cells

Electroporation was used to transform the binary vector system into *Agrobacterium*. The electroporator was set to 25 µF, 2.5 MV and 400 Ω. Plasmids and culture were maintained on ice. Cuvettes were stored in the freezer until imminently required. A 1 µl aliquot of the pSoup helper plasmid and 1 µl of the destination vector was added to 50 µl of *Agrobacterium* under a sterile flow hood. The pSoup plasmid was not included in transformation of the VIGS pTRV2 vector into strain GV3101, which already contained the pMP90 helper plasmid. The mixture was transferred to a pre-cooled cuvette on ice and slotted into the plastic mount of the electroporator. The mount was slid into the electrodes, electroporation was carried out and the cuvette was immediately returned to ice. It was ensured that the time constant on the electroporator read no lower than 9 ms to indicate sufficiently slow electroporation.

A 450 µl aliquot of LB was added to the transformed cells and transferred to an Eppendorf. This was incubated at 28 °C with shaking at 180 rpm until the end of the day, at which point they were plated on LB agar media containing appropriate antibiotics and returned to the incubator. A combination of kanamycin and rifampicin were each used at a concentration of 50 mg/ml to select for the pBRAC1 destination vector and pSoup helper plasmid, respectively. Gentamycin was also added at 50 mg/ml for selection of the pTRV2 VIGS vector.

2.14.3 Growth and Screening of Transformed *Agrobacterium*

The incubation of transformed *Agrobacterium* continued over approximately three nights until the selective plates presented growth. Four colonies were used to inoculate 10 ml of

LB containing kanamycin and rifampicin, each at a concentration of 50 mg/ml, which were incubated overnight at 28 °C with shaking at 180 rpm. A digestion had to be carried out to confirm final integrity of the transformed vectors. A miniprep was therefore used to obtain constructs from 8 ml of the cultures (Chapter 2.4.2) for transformation back into *E. coli* via heat shock (Chapter 2.12) because the higher copy number of *E. coli* over *Agrobacterium* provides greater material for digestion. The remaining *Agrobacterium* culture was used to form 30 % glycerol stocks for long term storage at -70 °C.

Positive *E. coli* colonies from the heat shock transformation were used to inoculate a liquid culture, which subsequently underwent miniprep the following day. The extracted destination vector underwent digestion with *BsrGI* (Chapter 2.11.11) to confirm integrity of the final plasmid in *Agrobacterium* before moving forward with plant transformation.

2.15 Transformation of *Primula vulgaris* Seedlings

The transformation of *P. vulgaris* was carried out to generate overexpression (Chapters 5.2 & 5.3) and knockdown (Chapter 5.4) mutants alongside plants with virally induced *KFB^T* silencing (Chapter 5.7) for the elucidation of its function. This was achieved via vacuum infiltration of seedlings with *Agrobacterium* cultures that were previously electroporated with the desired binary vector system. Dr Sadiye Hayta had developed a transformation system in *Primula* and continued to optimise the protocol throughout the duration of this PhD project (Hayta *et al.*, 2018). The final method is outlined here.

2.15.1 Vacuum Infiltration of *P. vulgaris* Seedlings

For vacuum infiltration of *Primula*, *P. vulgaris* seeds were sterilised (Chapter 2.2.1) and germinated on MS media (Chapter 2.3.1). They were grown beyond the cotyledon stage until their first mature leaves had formed. A 200 µl aliquot of transformed *Agrobacterium* (Chapter 2.14) was used to inoculate 10 ml of LB containing kanamycin and rifampicin at 25 mg/ml each. This was incubated overnight at 28 °C with shaking at 200 rpm. A 500 µl aliquot of this culture was used to further inoculate 50 ml of LB, supplemented with 5 µM of acetosyringone and the same antibiotic concentrations from the previous step. This was incubated overnight at 28 °C at 200 rpm.

The culture was decanted into a 50 ml falcon and centrifuged at 3200 g_n for 10 mins with the brake setting switched off. The supernatant was discarded and the pellet resuspended in 10 ml of $MgCl_2 \cdot 6H_2O$ infiltration buffer. Optical density of the sample was measured at 600 nm and the suspension was accordingly adjusted to 0.5 OD using the infiltration buffer. A 25 ml volume of this final culture was prepared to contain a 100 μM concentration of acetosyringone and incubated for 5 hrs at room temperature with shaking at 120 rpm.

Silwet L-77 was added to the infiltration culture at a concentration of 10 $\mu l/l$, using 4 μl from a 1 in 10 diluted stock solution. This is a commercial surfactant used here to aid in the uptake of *Agrobacterium* by reducing the surface tension of water. Approximately ten to fifteen seedlings were transferred into the 25 ml culture from the MS plates on which they had germinated. The open falcon containing the seedlings was positioned in a vacuum chamber and 600 mbar of negative pressure was applied. After 15 mins, the hose was pulled to rapidly release the vacuum before being reattached. This process was repeated a further four times across an hour. The tube was closed with a lid and left overnight at room temperature with shaking at ~ 120 rpm.

The transformed seedlings were washed six times in sterile distilled water and returned to the shaker for further overnight incubation in MS containing 250 $\mu g/ml$ of cefotaxime. Filter paper was used to dry the seedlings under sterile conditions before being planted onto MS media. They were allowed to recover for one week in a growth chamber before the tissue culture process was executed. Growth chamber conditions are provided in Chapter 2.1.

2.15.2 Tissue Culture of Transformed *P. vulgaris* Seedlings on Selective Media

After vacuum infiltration, the seedlings were chimeric for transformed and non-transformed cells (Hayta *et al.*, 2018). Tissue culture was used to induce callus from only transformed cells that were selected to eventually regenerate an entire mutant plant (Hayta *et al.*, 2019).

The following reagent stocks were made in advance: thidiazuron (TDZ) at 2 g/l, naphthaleneacetic acid (NAA) at 1 g/l, $AgNO_3$ at 1.7g/l, Timentin at 160 g/l and kanamycin at 125 g/l. The TDZ was dissolved in DMSO. The other reagents were dissolved in H_2O and filter sterilised with ~ 1.5 ml of NaOH used to help dissolve the NAA.

The callus induction media was made in two halves, which together amounted to 500 ml. The first half contained 2.5 g of agarose in 250 ml of H₂O and this was autoclaved, as described in Chapter 2.3. The second half contained 0.5 g of Gamborg B5 macro salts, 0.25 g of MS micro salt mixture, 0.052 g of MS vitamins and 15 g of maltose in 250 ml of H₂O. This underwent heating and stirring until the salts had dissolved. The pH of the mixture was adjusted to 5.8 before being filter sterilised into an autoclaved flask. It was maintained at 60 °C until use.

When the first half had completed the autoclave process, both medias were combined and mixed in a sterile laminar flow hood. They were left to cool below ~50 °C to protect the next reagents from degradation. When the mixture could be handled comfortably (and before the agarose had solidified), 250 µl of NAA was added with 500 µl each of TDZ, AgNO₃, Timentin and kanamycin from the aforementioned stock solutions. This was finally poured into petri dishes.

After a week of recovering on MS media, the transformed seedlings were cut into ~1 cm leaf sections and positioned with their adaxial surface on the *Primula* callus induction (PCI) media. The plates were sealed and returned to the growth chamber. Samples were transferred to fresh PCI media every two weeks. They were monitored for positive selection of transformed material by kanamycin and regularly checked for growth of contaminant fungi.

2.16 Transformation of *Arabidopsis thaliana*

While *Primula* would be the ideal target for transformation assays aimed toward discovering the true function of *KFB^T* in floral heteromorphy, transformation of *Arabidopsis* was also carried out due to its rapid life cycle and well-established protocols (Sanchez-Serrano & Salinas, 2014). The reliability of *Arabidopsis* transformation techniques has been proven and developed over decades, thus were used to gain initial clues regarding *KFB^T* function within a heterologous plant system.

2.16.1 Floral Dipping of *Arabidopsis* Plants

Floral dipping was used to transform *Arabidopsis* with vector constructs designed to generate *KFB^T* mutants (Figures 5.2 & 5.3). The Colombia Col-0 ecotype of *Arabidopsis*,

described in Chapter 2.1, was grown under long day conditions in trays with 3x3 rows of 7 cm pots containing 'Arabidopsis mix' by the John Innes Centre horticultural team. Growth was monitored until the optimal stage for transformation; when plants had many immature flower clusters and only a few fertilised siliques.

A 400 µl aliquot from a glycerol stock of *Agrobacterium* containing the desired construct was used to inoculate 10 ml of LB supplemented with kanamycin and rifampicin, both at 50 mg/ml. This was incubated overnight at 28 °C with shaking at 220 rpm. The culture was subsequently poured into 50 ml of LB containing the same antibiotic concentrations and returned to the incubation conditions overnight again.

The incubated culture was centrifuged for 10 mins at ~3200 g_n with the brake setting switched off. The supernatant was discarded and freshly made infiltration buffer was used to resuspend the pellet to an OD₆₀₀ of 0.8. At least 500 ml of infiltration medium was prepared for each construct to be transformed. A litre of infiltration medium contained 4.3 g of MS salts, 1x B5 vitamins, 5 % sucrose, 0.044 µM of 6-Benzylaminopurine and 0.03 % Silwet L-77. The pH was brought to 5.8 with KOH. The resuspended culture was kept cold on ice.

The *Arabidopsis* plants were watered prior to transformation, to minimise the uptake of infiltration medium by the soil. Any fertilised siliques on the plant were also removed with scissors to reduce the amount of unwanted non-transgenic seeds collected later. The culture resuspended in infiltration medium had been decanted into 400 ml plastic beakers. The plants were individually upturned and dipped into the culture for 15 secs, with their flowers and stems completely submerged in the beaker as the culture was gently agitated to ensure infiltration. Plants were immediately placed in a plastic biohazard bag and sealed for 24 hrs in the presence of light before being unbagged and allowed to continue growing and set seed.

2.16.2 Seed Collection and Transgenic Screening

Once transformed plants presented siliques, the watering routine was ceased and their aerial parts were bagged to capture any dispersed seeds. When the seed capsules had dehydrated and turned yellow, bags were squeezed to make the seeds fall to the bottom. The corner was cut off the seed bag and its contents were poured onto a sieve with a pore size of 300 µM. Sieving took place over a sheet of white A4 paper that had been creased

down the middle. Seeds were subsequently tipped down this crease into a 1.5 ml Eppendorf for storage.

The seeds were sterilised with chlorine gas (Chapter 2.2.2) and planted on MS media (Chapter 2.3.1) containing 50 mg/ml of kanamycin. The kanamycin resistance gene from the Gateway destination vector was co-transformed into the *Arabidopsis* genome alongside the *KFB^T* fragment of interest, therefore successfully transformed plants survived on the selective media and untransformed plants quickly turned colourless and died after germination. Genetically modified plants were kept in a CER chamber with a 16 h light period, a daytime temperature of 22 °C and a night temperature of 20 °C. Their seed was collected and second-generation plants were germinated and grown alongside wildtype *Arabidopsis* for fair comparison to identify mutant phenotypes.

2.16.3 Genotyping and Phenotyping Transgenic *Arabidopsis*

The Extract-N-Amp Plant PCR Kit from Sigma-Aldrich was used for genotyping second-generation plants and confirming successful transformation. A ~1 cm² section of young leaf tissue was cut from each plant, added to 100 µl of extraction solution and incubated in a 2 ml Eppendorf tube at 95 °C for 10 mins. A 100 µl aliquot of dilution solution was subsequently added to the incubated samples and 4 µl of DNA was used in a PCR with 10 µl of Extract-N-Amp ReadyMix, 4 µl of H₂O and 1 µl of each *nptII* primer (F: GAGGCTATTCGGCTATGACTGG; R: ATCGGGAGCGGCGATACCGTA) designed to the kanamycin selection gene. The PCR products were electrophoresed (Chapter 2.8.2) and those presenting a band were shown to contain the kanamycin resistance gene from a successfully transformed plasmid. Leaf samples and flowers across the full range of development were screened for mutant phenotypes (Figure 5.6).

2.17 Virus Induced Gene Silencing to Knockdown *KFB^T* Transcripts in *Primula*

To observe the functional effects of silencing *KFB^T* transcripts, a *KFB^T* VIGS construct was introduced into the Polyanthus horticultural variety of *Primula* (Figure 5.7-1). This variety was partially chosen due to their high commercial availability but mainly because their flowers all emerge from a central thick stem that is well-suited to injection, which *P. vulgaris* does not have. Open flowers, mature buds and dead leaves were removed prior to treatment. Plants were transferred into 1 l pots containing peat and sand compost and

moved to insect-free level 2 containment chambers at the John Innes Centre on the Norwich Research Park. The chamber was maintained at 15 °C during the day and 12 °C at night.

The *Agrobacterium* was prepared by using 400 µl from a stock containing the *KFB^T* VIGS vector to inoculate 10 ml of LB supplemented with kanamycin, rifampicin and gentamycin, each at a concentration of 25 mg/ml. A second inoculation was prepared with a negative control vector. This contained the vector backbone without anything at the cloning site and was used to prove that the VIGS machinery and injection process was itself not causing any mutant phenotype. The phytoene desaturase (PDS) gene was also used, the silencing of which leads to reduced chloroplast pigment and causes white patches on infected plant material. This was carried out as a positive control by Dr Jinhong Li to prove that the VIGS assay worked correctly. A pTRV1 stock was used to make two 10 ml LB inoculations. They were all incubated overnight at 28 °C with shaking at 200 rpm.

These overnight cultures were poured into 50 ml of LB supplemented with the same antibiotic concentrations and returned to the previous incubation conditions for a further overnight period. The infiltration buffer was prepared the following morning, containing 10 mM MgCl₂ and 10 mM MES. The cultures were centrifuged at ~3200 g_n with the brake setting switched off. The pellets were resuspended in 25 ml of infiltration buffer and the OD₆₀₀ was adjusted to 2.0 before the addition of 200 mM acetosyringone and subsequent incubation for 4-5 hrs at room temperature on a rocking platform (~40 rpm). The VIGS treatment was carried out immediately after this incubation.

Treatment was carried out in a level 2 containment glasshouse. Equal amounts of pTRV1 and pTRV2 cultures were combined and mixed within 50 ml Falcon tubes. A needle with a 0.6 mm pore was affixed to a 1 ml syringe for injection. After being filled with *Agrobacterium* mixture, the plunger was gently depressed while being slowly inserted into the stem of the plant until culture was no longer dripping from the needle tip and was instead being taken up by the stem. This was a good gauge of the optimum depth to insert the needle to find conductive vascular tissue without penetrating too far into (or entirely through) the plant stem. It was essential to very slowly depress the syringe and allow uptake of the medium at a rate manageable by the plant. In the best cases, droplets of *Agrobacterium* fluid could be seen emerging from the base of the floral umbel, proving that the viral medium had been transported to the flowers and the vascular tissue was loaded to maximum capacity.

Each stem was injected a single time until maximum uptake, which depended on the stem maturity and volume. If more than one wound occurred – either by accidentally pushing the needle all the way through the stem or by making multiple penetration attempts – then a path of least resistance was created and acted as a plughole that the medium simply flowed through instead of reaching the apical flowers. Likewise, if the syringe plunger was overly forced, the stem would entirely rupture and become unusable for VIGS.

It was important to wear safety goggles because the plant can resist injection with such pressure to spray the medium back upon removal of the needle. Treatment was repeated weekly until each stem had been treated twice and the plant had stopped presenting new flower buds. Young stems that were too thin and flexible to properly treat were left to mature until deemed prime for injection during one of the later weekly repeats. Plants were regularly monitored and any mutant phenotypes were photographed. Mature flowers were also self-crossed to check for modifications of the self-incompatibility system.

2.18 Execution of the *GUS* Reporter Gene Assay

To visualise *in situ* location of KFB^T , a construct had been assembled containing the KFB^T promoter and 5'-UTR regions driving expression of the *GUS* reporter gene (Figure 4.8-1). It was electroporated into *Agrobacterium* (Chapter 2.14) and transformed into *A. thaliana* (Chapter 2.16) due to its rapid life cycle and reliable protocols. The *GUS* assay solution was made in two parts. Containers at all stages were wrapped in foil to exclude light. The first part consisted of 70 mg of X-gluc (5-Bromo-4-chloro-3-indolyl- β -D-glucoronide) added to 2 ml of dimethyl sulphoxide (DMSO). For the second solution, 7.602 g of sodium phosphate was added to 150 ml of H₂O. The pH was adjusted to 7.0 with concentrated hydrochloric acid before the addition of 4 ml of Na₂EDTA (0.5 M) and 200 μ l of Triton X-100 – the latter should be pipetted with a severed P1000 tip due to its viscosity. Both solutions were mixed, diluted to 200 ml and stored at -20 °C in 50 ml aliquots.

Working solutions were maintained at 4 °C. Leaves, root samples, siliques and flowers from all developmental stages of transgenic *Arabidopsis* plants were excised and added to the thawed *GUS* mix before incubation at 37 °C with shaking at ~70 rpm. Samples were washed in 100 % ethanol and incubated overnight in fresh ethanol on a rocking platform at ~45 rpm to remove chlorophyll. This was subsequently replaced with 50 % ethanol before inspection under a light microscope for blue *GUS* activity.

2.19 Timing and Location of KFB^T Protein Expression

A series of qPCR experiments (Chapters 4.2-4.4) were designed to investigate the spatial and temporal expression of *KFB^T* at the transcript level. Antibodies were raised against *KFB^T* to also carry out these investigations at the protein level. This was designed to acquire data that either supported the qPCR findings or provided alternative models. For example, differences between the two datasets could be indicative of post-transcriptional regulation or protein accumulation that highlight important areas of the *KFB^T* pathway and its function in floral heteromorphy. Flower buds had been preserved and sectioned in preparation for immunolocalisation assays. A dot blot was therefore utilised to identify a suitable antibody and Western blots were used to locate *KFB^T* in the dissected whorls of *P. vulgaris* flowers across several developmental stages.

2.19.1 Production of KFB^T Antibodies

Specific peptides had been previously designed to generate polyclonal antibodies for the five *S* locus genes (Kent, 2016). The *KFB^T* peptide sequence used was: EVIPGLPEDLGLE. These antibodies were produced by Dundee Cell Products in two rabbits. They were injected four times each and immunoaffinity chromatography was used to isolate the antibodies from antisera after the second, third and final immunisations.

2.19.2 Identifying a Suitable KFB^T Antibody

An initial dot blot was carried out to confirm interaction between the *KFB^T* peptide and antibody (Chapter 4.6). Before testing affinity with *KFB^T* extracted from *Primula*, interaction was first tested against the peptide that the antibodies were raised against. Dundee Cell Products synthesised two peptide conjugates; one bound to bovine serum albumin (BSA) and a second conjugated to keyhole limpet hemocyanin (KLH). They provided images of their affinity purification that showed the most concentrated antibody fractions. The aim of this initial dot blot was to show interaction with both peptide samples (thus confirming antibody affinity with *KFB^T* and not just the BSA or KLH tags).

The peptide conjugates were at a concentration of 2 mg/ml and 2 µl of each were dotted onto separate nylon membranes. These were blocked in 5 % milk for 90 mins. Incubation steps were carried out at room temperature on a rocking platform. The blocking solution was made from milk powder dissolved in phosphate-buffered saline (PBS). This process

allowed milk proteins to fill any vacant area around the sample, ensuring the antibody could only bind to the protein of interest and any excess was washed away. This is because the membrane has a very high affinity for any protein and would otherwise simply bind to the antibody itself and give unusably high background results and present false positives. Milk was used as a blocking substrate to easily obtain unrelated proteins that were assumed to exhibit no binding activity with the KFB^T antibody.

The membranes were further incubated for 2 hrs in a 50 ml solution containing the primary antibody at a dilution of 1 in 10,000. The membranes were subsequently washed in 1x PBS for 10 mins. The PBS was replaced and this was repeated for 20 mins and finally for 40 mins to wash away the primary antibody, which refers to the antibody raised against KFB^T. The secondary antibody was an anti-rabbit immunoglobulin G from goat, which binds to the primary antibody. It was conjugated to horseradish peroxidase by the manufacturer to undergo a light-emitting chemical reaction that allowed for visualisation of the target protein to confirm presence and successful binding. The membranes were incubated overnight in a 25 ml solution of 5 % milk containing 0.5 µl of this secondary antibody (2 mg/ml).

Substrates from the SuperSignal West Femto Chemiluminescent kit were used to trigger the horseradish peroxidase reaction. The peroxide buffer and luminol enhancer were mixed at a 1:1 ratio and 800 µl was added to each membrane, which were positioned on transparent laminate plastic. A second sheet of transparent plastic was laid over the top and air bubbles were removed before touching the edge against blue roll to absorb any excess drops. The membranes were immediately enclosed in a light-proof box for transport to the photographic developer machine. In a darkroom, sheets of photographic film were laid over the membranes for different time lengths to provide various exposures.

2.19.3 Crude Protein Extraction from *P. vulgaris* and Concentration Normalisation

The KFB^T antibodies had been preliminarily screened for affinity to the peptide that they were raised against. Before carrying out an immunolocalisation experiment to visualise the location of KFB^T *in situ*, the antibody was initially tested for affinity with denatured KFB^T within a crude protein pool extracted from *P. vulgaris* (Chapter 4.6). Concentrations of extracted protein pools had to be normalised across the sample set prior to this preliminary affinity test.

Flower samples were previously excised from petioles and stored in a -70 °C freezer. For protein extraction, samples were ground by hand with a mini pestle in a 1.5 ml Eppendorf containing 2x LDS buffer. The buffer was incrementally added as required until the tissue was ground into a viscous solution; adding too much leads to an overly diluted protein sample and too little gets entirely soaked up by the bud. A 2x LDS concentration was used and was estimated to dilute into a 1x solution due to the high water content already present in the flower tissue. The ground samples were boiled at 80 °C for 10 mins.

A gel was used to estimate the amount of extracted protein so that concentrations could be normalised across the sample range before use in a Western Blot assay. The samples were briefly centrifuged to pellet any tissue debris to the bottom of the tube. To prepare them for loading, a 5 µl aliquot of each protein extract was used to make a 20 µl solution containing 10 % dithiothreitol (DTT; 0.5 M), 5 µl of H₂O and 10 µl 2x LDS. The samples were boiled in this solution for a further 10 mins at 80 °C.

To run the protein gel, 950 ml of H₂O was added to 50 ml of 20x NuPAGE™ SDS Running Buffer (ThermoFisher NP0001). The packaging, tape and comb was removed from a premade gel before having its plastic cassette and wells briefly rinsed in H₂O. It was clamped vertically into the electrophoresis apparatus and an empty cassette was fitted into the second slot, if only a single gel was being run. A 500 µl aliquot of NuPAGE™ antioxidant (ThermoFisher NP0005) was added to 200 ml of the running buffer (immediately prior to running the gel) and poured into the central compartment between the two clamped cassettes. Upon ensuring no leaks were occurring between the apparatus partitions, the outer compartment was filled with the remainder of the running buffer that did not contain antioxidant. The boiled samples were pipetted into the wells alongside an appropriate protein ladder and 200 V was applied for 50 mins.

After electrophoresis, the fragile gels were carefully disassembled from their plastic cassettes and incubated in InstantBlue Ultrafast Protein Stain (Sigma-Aldrich ISB1L) for 1 hr on a rocking platform. The intensity of bands was judged on the gel lanes and necessary adjustments were estimated regarding the volume of each protein extract to be used in subsequent analyses that were required to achieve equal concentrations across the sample range.

2.19.4 Electrophoresis Transfer of Protein Bands from Gel to PVDF Membrane

After the initial loading gel, a second gel repeat was run containing normalised concentrations across the samples (Figure 4.6-1). The vertically separated bands then had to be horizontally transferred onto polyvinylidene fluoride (PVDF) membrane to immobilise the proteins for downstream Western Blot analysis. To transfer one gel, 1 ml of NuPAGE™ Antioxidant was added to 50 ml of NuPAGE™ Transfer Buffer and 100 ml of methanol before being diluted with H₂O to a final volume of 1 l. To transfer two gels simultaneously, the same solution was made containing 20 % methanol instead.

Prior to carrying out the gel transfer, this buffer solution was used to thoroughly soak six blotting pads until their air bubbles were completely removed. PVDF membrane and two sheets of filter paper were all cut to the dimensions of the gel. The work area must be clean around exposed PVDF membrane to avoid contamination due to its high affinity for proteins. It was soaked in methanol for 30 secs, briefly rinsed in deionised water and kept in the transfer buffer for several minutes before use.

The cassette was disassembled and wells were excised from the gel. Filter paper was briefly soaked in transfer buffer and laid just above the lip at the gel foot, with air bubbles being removed. The gel was overturned, the foot was cut away and the surface was wetted with transfer buffer before being overlaid with PVDF membrane. Air bubbles were removed and the second piece of filter paper was placed on top. Guidance for assembling the apparatus for two gels is provided in the NuPAGE™ Bis-Tris Gel Instruction Booklet by Invitrogen Life Technologies.

The gel sandwich was placed onto two of the pre-soaked blotting pads inside the cathode core. The remaining space was filled with blotting pads until they rose 0.5 cm above the rim of the core before clamping it shut inside the electrophoresis block. The central core was filled with transfer buffer. The outer space was filled with deionised water until it reached 2 cm below the rim of the apparatus. This acts as coolant during electrophoresis, which was carried out at 30 V for 1 hr.

2.19.5 Western Blot Screen of KFB^T in Crude Protein Extracts from *P. vulgaris* Flowers

A Western Blot had been used to identify which rabbit had produced antibodies exhibiting greatest affinity for KFB^T (Figure 4.6-1). However, eight fractions of this antibody were available from the blood purification step carried out by Dundee Cell Products. A Western

Blot was therefore required to identify the most suitable fraction for use in downstream immunolocalisation assays (Figure 4.6-2). This was carried out on the denatured crude protein extract from ~5 mm stage one *P. vulgaris* flower buds (defined in Chapter 2.1.1). In addition to checking validity of the antibody, this assay also confirmed its affinity for denatured KFB^T protein from the plant – as opposed to the synthesised peptide conjugates previously screened.

The protein pools extracted from *P. vulgaris* were separated vertically by electrophoresis (Chapter 2.19.3) and then horizontally transferred to PVDF membrane (Chapter 2.19.4). The transfer apparatus was disassembled and rinsed in PBS. The PVDF membrane was blocked for 20 mins in 5 % milk, as in Chapter 4.6.2. The milk was replaced and the primary KFB^T antibody was added. Antibody dilutions of 1:1000 and 1:10,000 were trialled. The immobilised proteins were incubated in this solution for 2 hrs on a rocking platform.

The PVDF membrane underwent three PBS washes for 10 mins each before three further washes in PBS containing 0.1% Tween. The blocking process was repeated for at least 15 mins. The secondary antibody was subsequently added directly to the milk and maintained in a 4 °C cold room overnight on a rocking platform. Secondary antibody dilutions of 1:100,000 and 1:1000,000 were trialled. This goat antibody was an anti-rabbit horseradish peroxidase conjugate. The previous PBS and PBS-Tween washes were repeated. In accordance with Chapter 2.19.2, chemiluminescent substrates were added and photographic film was developed to visualise presence of any KFB^T protein on the PVDF membrane.

A staining solution was prepared that contained 250 ml of H₂O, 200 ml of methanol, 50 ml of acetic acid and 0.5 ml of Coomassie Blue. The Western Blot was stained for 5 secs in this solution to reveal the proteins from the crude *P. vulgaris* flower extract that were previously size-separated. The staining solution was also made without the Coomassie Blue and used to wash the blot three times at 10 mins each. The translucent developed photographic film was then laid over the stained membrane to align blots against corresponding protein bands, to allow protein size and sample lane of blots to be identified.

2.20 Yeast Two-Hybrid

Yeast two-hybrid assays were used to test for protein-protein interactions involving KFB^T (Chapter 6). This could identify the protein targeted for degradation by KFB^T and offer functional insight. The paired mating assay (Chapter 2.20.1) was used when the target protein to be tested was known, otherwise protein libraries were screened when this partner was unknown (Chapter 2.20.3).

2.20.1 Paired Individual Mating

The interaction between KFB^T and a range of known *P. vulgaris* proteins were tested via yeast two-hybrid mating. The KFB^T coding region had already been assembled into the pGADT7 and pGBKT7 activation and binding domain vectors (Figure 6.1). Binding and activation domain vectors are respectively referred to as bait and prey constructs. Bait constructs were transformed into the Y187 yeast strain and prey constructs were transformed into AH103 cells (Chapter 2.13). These two cell types are compatible for mating (Matchmaker™, 2007).

A GLO^T fragment had been prepared in both bait and prey constructs by Dr Barry Causier via Gateway cloning for a previous study at the University of Leeds. Two further genes, *PvGLO* and *PvDEF*, were also prepared in the same way by Dr Causier. These proteins are known to interact with each other (Tröbner *et al.*, 1992) and were therefore used as a positive control to ensure the mating and yeast two-hybrid screening functioned correctly. Stocks of KFB^T , GLO^T , *PvDEF* and *PvGLO* were therefore obtained as both bait and prey constructs transformed (Chapter 2.13) into yeast strains Y187 and AH103, respectively.

A 10 µl aliquot from each prey sample was pipetted onto YPDA general purpose growth media (Chapter 2.3.3) and 10 µl aliquots of the KFB^T bait sample were pipetted on top. To simultaneously execute this test in the reverse direction, the KFB^T prey sample was also pipetted onto aliquots of each bait sample too. Mating assays were therefore carried out between KFB^T against GLO^T , *PvDEF*, *PvGLO* and itself. Samples of the known interacting pair, *PvDEF* and *PvGLO*, were mated with each other as a positive control. The *PvGLO* bait and prey constructs were also mated with each other, which was a suitable negative control because they are known not to homodimerize. All plates were incubated at 30 °C for three days to allow mating between the compatible cells to take place.

The centre of each sample of mated cells was swabbed with a pipette tip and resuspended in 20 µl of sterile TE buffer before being transferred to -WL SD media plates lacking tryptophan and leucine (Chapter 2.3.4). These plates selected only for diploid cells that had successfully mated to contain the tryptophan and leucine biosynthesis genes on the pGBKT7 and pGADT7 plasmids, respectively. The plates were incubated at 30 °C for a further three days before this suspension and transferral procedure was repeated onto -WHL media.

The -WHL plates lacked tryptophan (W), leucine (L) and histidine (H). This media selected for diploid cells containing constructs that expressed a pair of interacting proteins that bring together the binding and activation domains of the respective fusion proteins to enable expression of the histidine biosynthesis gene. Cells without interacting partner proteins cannot survive in the absence of histidine. The test was repeated with the addition of 3-amino triazole (3-AT) to the -WHL media at concentrations of 2.5, 5, 7.5 and 10 mM. This molecule inhibits binding and indicated affinity between the interacting proteins. Dr Barry Causier (University of Leeds) had previously confirmed a minimum 3-AT concentration of 2.5 mM sufficiently prevented auto-activation of the *KFB^T* construct (Figure 2.3.4). The results of the screen were observed after five days (Figure 6.2-1).

2.20.2 Obtaining *Primula vulgaris* and *Arabidopsis thaliana* Y2H Libraries

Yeast two-hybrid libraries of *Arabidopsis* and *Primula* were used to screen for proteins that interact with *KFB^T* (Chapter 6). RNA was extracted from *Primula* flower buds by Dr Jinhong Li and the Matchmaker Library Construction Kit was subsequently used by Dr Barry Causier at the University of Leeds. This involved synthesising cDNA from the RNA pool by using an oligodT primer with a 3'-end adaptor sequence. A second adaptor sequence was attached to the 5'-end of the cDNA fragments during synthesis. Primers designed against these adaptor sequences were used to amplify the cDNA library via PCR.

Amplified products were co-transformed into the AH103 *Saccharomyces cerevisiae* yeast strain alongside the pGADT7 prey vector, which also contains the tag sequences. The cDNA fragments were thereby incorporated into pGADT7 via homologous recombination that occurred between the cDNA adaptors and these matching vector sequences. This resulted in a library of AH103 yeast cells that each produced a protein from a *Primula* cDNA fragment fused to the activation domain of the vector.

The process was repeated to generate the *Arabidopsis* library and an aliquot was sent to the Gilmartin lab. Approximately 1.5×10^7 cells were plated across one hundred 15 cm plates of -L SD media and incubated for 4 days. Colonies were subsequently scraped from the plates into YPDA liquid containing 25 % glycerol for long-term freezer storage. A thawed aliquot underwent serial dilution before spreading onto plates for further incubation to calculate the viable cell density of the library.

2.20.3 Yeast Two-Hybrid Protein Library Screens

Protein libraries were screened via yeast two-hybrid to identify potential partners of KFB^T (Chapter 6). Medias were prepared in advance and the entire screening procedure required 50 ml of -W gel, 50 ml of -W liquid, 50 ml of 2x YPDA liquid, 10.5 ml of 0.5x YPDA liquid, 200 ml of -WL gel and 1 l of -WHL gel (Chapters 2.3.3 & 2.3.4). The KFB^T -DNA binding domain vector had been transformed into *Saccharomyces cerevisiae* strain Y187 (Chapter 2.13), which is compatible for mating with the AH103 strain that was made to contain the cDNA library in the activation domain vector.

A Y187 colony transformed with the KFB^T bait plasmid (Figure 6.1 upper) was selected from a -W plate and used to inoculate 55 ml of -W liquid SD media. This was incubated at 30 °C for three nights and a 1 ml aliquot was used to make a glycerol stock for long term freezer storage, 4 ml was used in a miniprep (Chapter 2.13) to confirm vector integrity via subsequent digestion (Chapter 2.11.11) and the remaining 50 ml was carried forward into the mating procedure for a yeast library screen.

A 100 μ l volume of culture was used with a haemocytometer under a light microscope to estimate average cell density. This was used to calculate a volume containing 5×10^8 bait cells to be mixed with 2.5×10^8 AH103 cells from a pre-enumerated *Primula* or *Arabidopsis* prey library. The cell mixture was centrifuged at 1000 g for 10 mins and resuspended in 50 ml of 2x YPDA liquid media. The suspension was incubated in a 2 l flask across three nights at 30 °C with shaking at 50 rpm. It was essential that shaking did not exceed this speed or mating does not take place.

The mated cell suspension was centrifuged at 1000 g for 10 mins and the pellet was resuspended in 10.5 ml of 0.5x YPDA liquid media. A 10 μ l aliquot of the culture was used to plate serial dilutions of 10 thousand, 100 thousand, 1 million and 10 million on -WL plates to select for diploid cells containing both the library activation domain and KFB^T binding

domain plasmids. This screened for successfully mated yeast. It was not necessarily suggestive of protein partner interactions but the number of growing colonies was instead used to estimate the total number of genes assayed by the yeast two-hybrid library screen.

The remaining culture was spread in 500 μ l aliquots across 20 -WHL SD plates containing 2.5 mM of 3-AT. This media selected only for diploid cells containing a pair of interacting proteins. All yeast plates were incubated at 30 °C. Colonies on the -WL plates could be enumerated after approximately four days to calculate the total number of genes screened in the yeast two-hybrid library. After two to three weeks, positive colonies on the -WHL media were each transferred to 10 μ l of TE buffer, which was then pipetted onto fresh -WHL media (but not spread). This allowed a larger circular colony to grow from the droplet, which was used to inoculate 6 ml of -WHL liquid media before incubation at 30 °C with shaking at 180 rpm.

The incubated samples subsequently underwent a miniprep a procedure (Chapter 2.4.3) and PCR was carried out on the extracted DNA using primers designed to either side of the cDNA insert within the activation domain vector. The library fragment was therefore sequenced so the interacting partner protein could be identified. A 50 μ l PCR volume was used containing 10 μ l of Phusion High-Fidelity buffer, 1 μ l of dNTPs, 0.5 μ l of Phusion polymerase, 2 μ l of crude DNA extract template from the yeast miniprep and 2.5 μ l of each primer (F: TCATCGGAAGAGAGTAG; R: GTGAACTGCGGGTTTTTCAGTATCTACGAT). The reaction was heated to 98 °C for 1 minute before undergoing 35 cycles of 98 °C for 30 secs followed by 60 °C for 30 secs and 72 °C for 3 mins. After a final 72 °C extension period of 5 mins, 10 μ l of PCR product was electrophoresed on a gel to confirm only one band was present in each lane. Samples were directly purified (Chapter 2.8.2) and sequenced by Eurofins Genomics to identify the cDNA library fragment that produced the interacting partner protein.

Chapter 3

3 Bioinformatical Analysis of the KFB^T

The KFB^T gene was initially identified by Dr Jinhong Li while locating the *Primula vulgaris* S locus (Li *et al.*, 2016). It was found to be homologous to the four *Kiss Me Deadly* genes in *Arabidopsis thaliana*. The Arabidopsis Information Resource (TAIR; www.arabidopsis.org) indicated that these genes were also known as *Kelch Repeat F-Box* or *Kelch-Domain-Containing F-Box* and so the *P. vulgaris* gene was thereafter called *Kelch F-Box^{THRUM}* (KFB^T).

The KFB^T sequence was analysed to gain early information about its function that aided in forming initial hypotheses and directing the lab work. This involved sequence alignments to homologous genes (Chapter 3.4), transcript quantification (Chapters 3.7 & 3.8), finding predicted regulatory elements in the 5'-promoter region (Chapter 3.6), identifying functional domains in the coding sequence (Chapters 3.1 & 3.5), finding other genes in *P. vulgaris* that contained these domains (Chapter 3.2) and investigating their phylogeny. The KFB^T gene was also identified in other members of the *Primulaceae* family as part of a study to explore conservation of the S locus across species (Chapter 3.3).

Bespoke programs and scripts were written personally for the project. This included tools for dot matrix alignment (Chapter 3.6), nucleotide translation (Chapter 3.3), stem-loop structure screening (Chapter 3.6), comparative genomics (Chapters 3.3 & 3.4), non-synonymous mutation enumeration (Chapter 3.3) and Kelch motif queries (Appendix A). Notable exceptions were the Clustal (Larkin *et al.*, 2007; Chapters 3.1-3.4) and HISAT (Kim *et al.*, 2015; Chapters 3.6 & 3.7) alignment packages as well as the StringTie (Pertea *et al.*, 2016; Chapters 3.7 & 3.8) assembler.

Bioinformatics was carried out in the Bourne Again Shell (www.gnu.org/software/bash), using its in-built command library on the cluster at the Earlham Institute, developed on a Red Hat Linux kernel and maintained by the Computing Infrastructure for Science department at the Norwich Research Park. Graphs were generated in version 3.4.3 of R (www.r-project.org). Sequence alignments carried out in the command line interface used version 2.1 of ClustalW (Larkin *et al.*, 2007) and those executed in the web interface utilised the online version of Clustal Omega (Sievers *et al.*, 2011).

3.1 Location of the Kelch Domain in KFB^T

The Kelch domain consists of a number of tandem motif repeats that each form one blade of a β -barrel propeller for directly aiding protein-protein interactions (Li *et al.*, 2004). There are usually five to seven Kelch repeats but plants often have fewer (Schumann *et al.*, 2011). Kelch motifs are variable and may only loosely obey a consensus sequence with as little as 11 % similarity (Bork & Doolittle, 1994).

To locate the Kelch domain in KFB^T and identify the number of Kelch repeats within it, Clustal Omega (Sievers *et al.*, 2011) was used to align the KFB^T amino acid sequence to the following Kelch consensus sequence: PRSGAGVVVVGGKIYVIGGFDGSQSLSSVEVYDPETNTWEKLP SMP (Prag & Adams, 2003). Upon aligning this consensus sequence to the best match in KFB^T, the identified region was manually removed from the KFB^T query sequence and the alignment was repeated to find the next closest match. This process continued until no further reasonable matches could be found.

A scoring system was established to rank the similarity of the identified Kelch repeats. Amino acids that perfectly matched the consensus sequence gained three points, those that shared highly similar properties scored two points, those that shared lowly similar properties were granted one point and totally dissimilar amino acids received no points. A sequence that perfectly matched the entire 46-amino acid consensus sequence would have achieved 138 points. The score for each Kelch repeat motif found in KFB^T was converted into a percentage of that highest possible score.

Five Kelch motifs were confirmed in KFB^T, with the fourth and fifth repeats intervened by 29 amino acids containing a fractional Kelch motif (Figure 3.1). The other four repeats were almost entirely continuous.

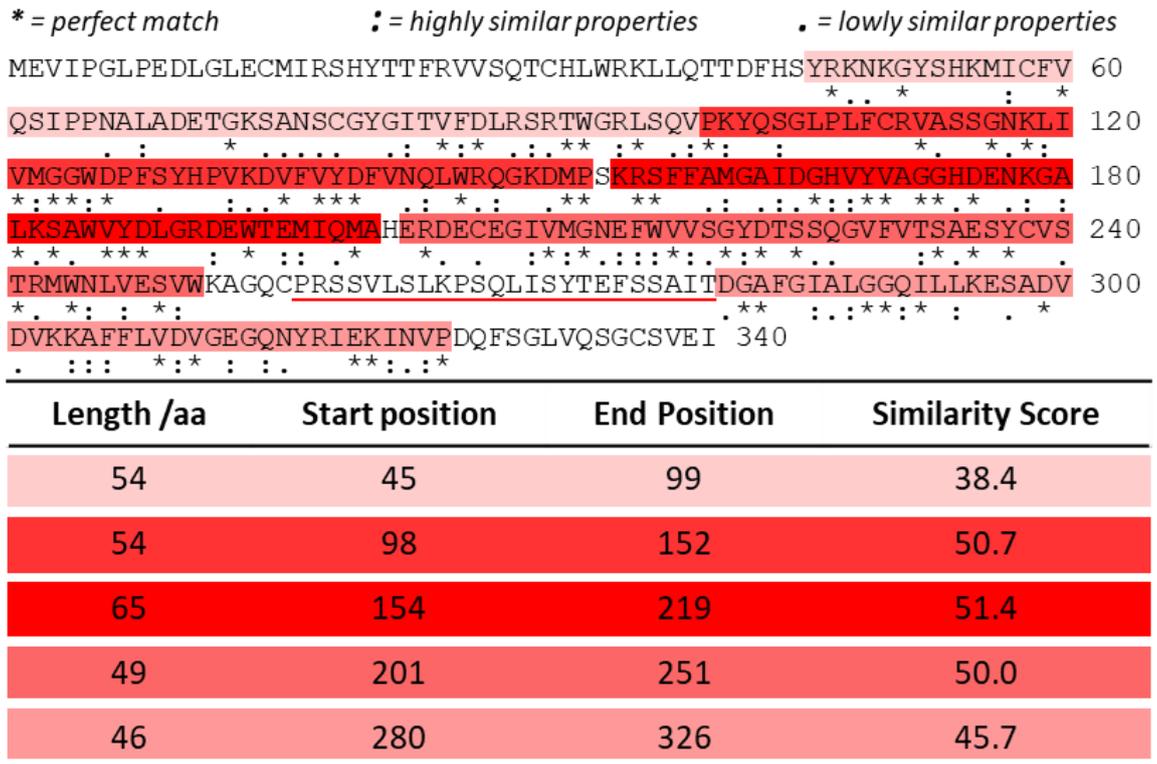


Figure 3.1: Five Kelch repeats were identified in KFB^T (upper). The underlined region represents identification of a fractional motif. Similarity of each motif to the Kelch consensus sequence (Prag & Adams, 2003) is shown (lower).

3.2 Identification of Kelch Proteins Encoded by the *P. vulgaris* Genome

Amino acid sequences derived from the *P. vulgaris* genome were screened for other potential members of the Kelch protein family to explore the possibility that other proteins related to KFB^T may also exist in *P. vulgaris*. Phylogenetic analysis of these candidates was carried out to potentially identify a protein that KFB^T had derived from or, if the function of its most closely related proteins was known, offer insight into the possible role of KFB^T in floral heteromorphy.

The genome of the long homostyle mutant was used for this investigation (Cocker *et al.*, 2018). It was assembled by Dr Jonathan Cocker from the DNA of a single plant and is homozygous for the *S* locus, therefore making it the best draft genome available for *P. vulgaris*. The only known mutation in this genome exists in the *CYP^T* gene (Li *et al.*, 2016), the dysfunction of which leads to reduced suppression of style length usually observed in thrums and therefore causes the long homostyle phenotype instead.

The genome had been annotated by Dr Jonathan Cocker, which provided the general feature format (GFF) file that was here used to acquire the coding sequences for each gene. A script was written (Appendix A1) that utilised coordinates from the GFF to locate exon positions of every *P. vulgaris* gene and join them together, thereby obtaining the complete coding sequence for every gene in the genome. Of the 24599 genes presented by the GFF, 15 % (3812 genes) were discarded due to not having an appropriate codon at either the start or stop position. This is likely due to the automated annotation process detecting RNA reads that map to areas without complete sequence data available.

The remaining genes were translated into amino acid sequences. The C program written to achieve this is available on GitHub (github.com/calumraine/cranslate). The following grep query was used in the Linux command line to search the amino acid sequences from each gene: `grep -B1 -E GG.{9,25}[LFY].....W.*GG.{9,25}[LFY].....W` (Figure 3.2-1). The grep command is a tool used to search files and extract lines containing at least one match to a given query. The -E flag was used to dictate that the search term was in the format of an extended regular expression and the -B1 flag requests for both the matching line and its preceding line to be printed, which here meant gene names would be obtained from the FASTA file as well as the matching sequences.

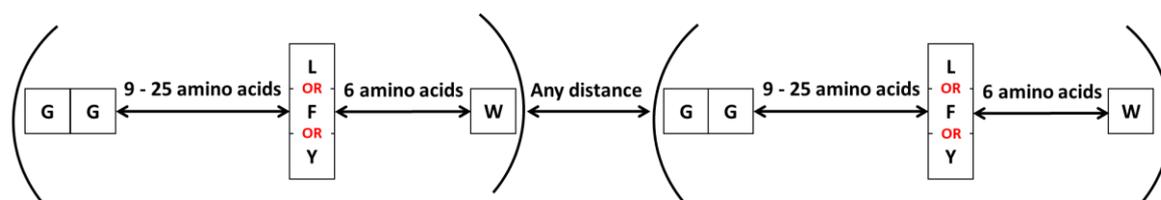


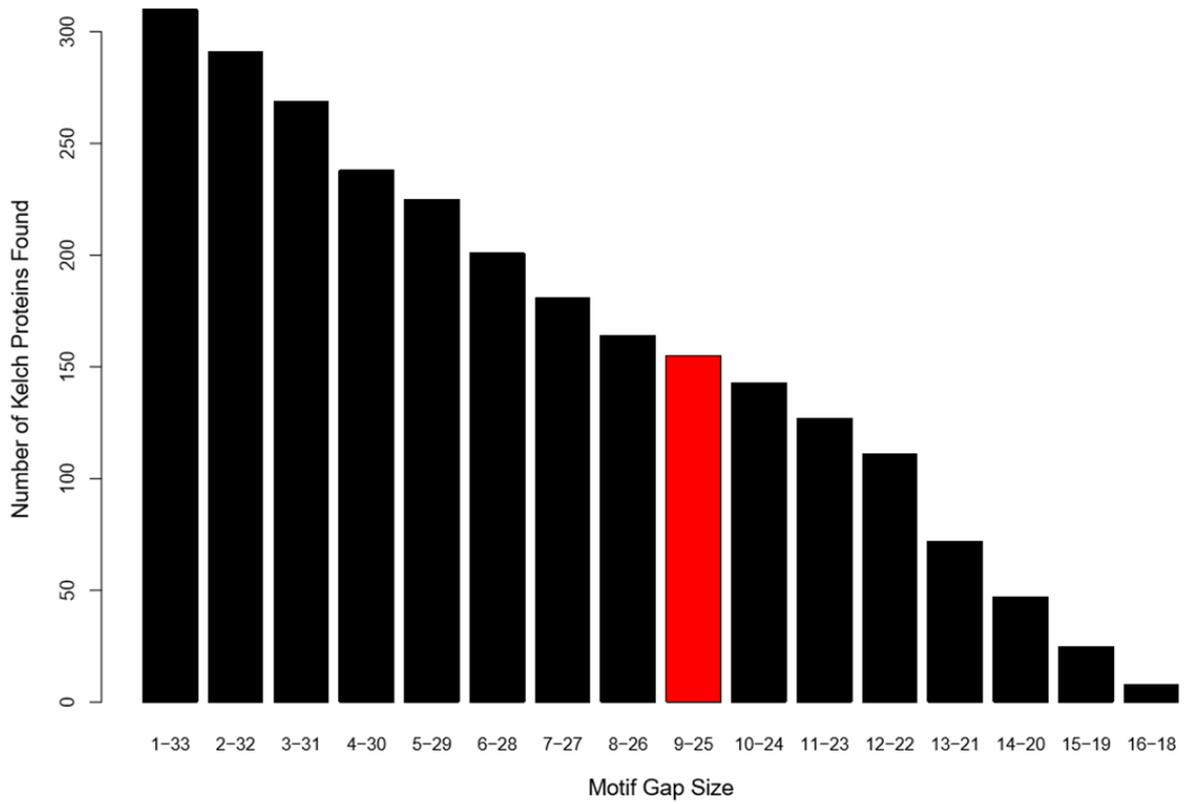
Figure 3.2-1: Key motif parameters and amino acid landmarks used to identify Kelch repeat proteins in *P. vulgaris*.

This query searched for the glycine pair (a fundamental signature of the Kelch consensus sequence) followed by between nine and twenty-five amino acids of any identity – a number estimated from literature review (Prag & Adams, 2003; Andrade *et al.*, 2001; Adams *et al.*, 2000). The query parameters ensured this linker was immediately followed by a commonly occurring leucine (hydrophobic) or phenylalanine (hydrophobic) or tyrosine (aromatic) residue and a final signature tryptophan separated by a linker containing exactly six of any amino acid. The length of this linker was used due to its occurrence in ~70 % of analysed Kelch domains (Adams *et al.*, 2000). The query ignored candidates that did not contain at least two of these Kelch repeat motifs separated by any number of amino acids. The search was repeated with modified sizes of the 9 to 25 amino acid linker to see how

this affected the number of results. Amino acid sequences from candidate genes identified by this Kelch motif screen were aligned against KFB^T (Figure 3.5 lower; Appendix A2) to compare similarities and begin phylogenetic analysis.

This motif query identified 155 genes in the *P. vulgaris* genome that encode proteins with a Kelch repeat domain (Figure 3.2-2 upper). One of these candidates was KFB^T, which provided reassurance that the search technique had worked. Each identified Kelch protein was aligned to KFB^T and the number of matches was converted into a percentage of the total 340 amino acid length of KFB^T (Figure 3.2-2 lower). Their similarities ranged from 7.65 % to 44.71 %.

Varying Search Parameters and the Number of Kelch Proteins Identified



Similarity Between KFB^T and Other *P. vulgaris* Kelch Proteins

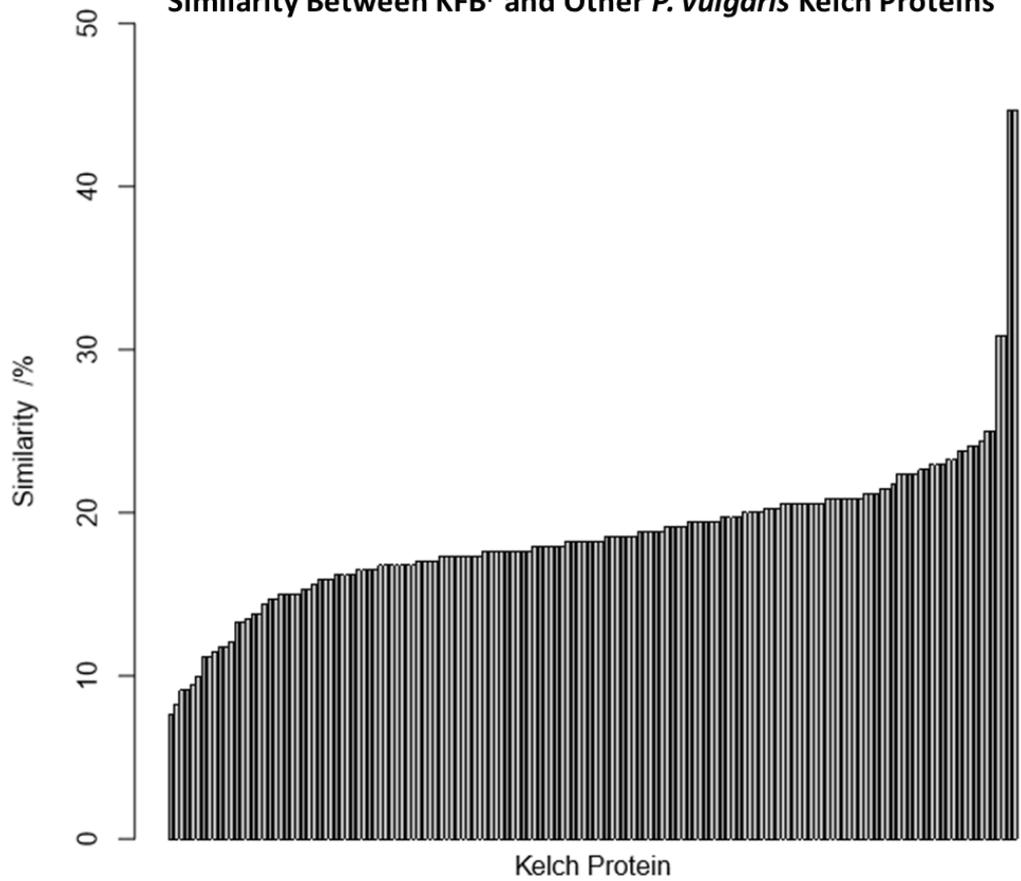


Figure 3.2-2: Kelch proteins identified in *P. vulgaris*. The Kelch motif consensus sequence contains a linker region of variable length. Several ranges for this linker were tested (upper) but a value of 9 to 25 amino acids was used to identify 155 Kelch proteins. Their similarity to KFB^T was calculated (lower).

A BLASTp search was carried out on amino acid sequences of the top fifteen Kelch proteins (Figure 3.2-2 lower) encoded by *P. vulgaris* that presented greatest similarity to KFB^T (Table 3.2). One of them was exactly the same length as KFB^T and presented a top match to KMD3. The third most similar *P. vulgaris* Kelch protein to KFB^T was found to match a KMD2-like protein, seven more identified as Kelch F-box proteins, a further Kelch protein was identified and a final match was found to have involvement in the ubiquitinase degradation pathway.

Table 3.2: Top BLASTp results for fifteen Kelch proteins from *P. vulgaris* that demonstrated greatest similarity to KFB^T. The number of matching amino acids was enumerated and converted to a percentage of the total query length.

Rank	Coding Sequence Length /aa	Similarity to KFB ^T Amino Acid Sequence		Top BLAST Hit
		Matches	%	
1	340	152	44.7	Kiss Me Deadly 3 like
2	349	152	43.55	F-box/Kelch-repeat protein
3	352	105	29.83	Kiss Me Deadly 2 like protein
4	356	105	29.49	F-box/Kelch-repeat protein
5	1785	85	4.76	Arginine-tRNA ligase
6	439	85	19.36	F-box/Kelch repeat protein
7	359	83	23.12	F-box/Kelch repeat protein
8	359	82	22.84	F-box/Kelch repeat protein
9	965	82	8.50	Kelch-like protein
10	357	81	22.69	Ubiquitin carboxyl-terminal hydrolase like
11	1226	81	6.61	RNA-binding protein Musashi Rbp6 like
12	737	79	10.72	F-box/Kelch-repeat protein
13	413	79	19.13	F-box/Kelch-repeat protein
14	870	78	8.97	Rab9 effector protein with Kelch motif like
15	667	78	11.69	Acyl-CoA-binding domain-containing protein

Alongside the similarity assessments of these Kelch proteins, Clustal Omega (Sievers *et al.*, 2011) was used to generate a phylogenetic tree of all 155 candidates. The clade containing KFB^T is displayed in Figure 3.2-3. Contrary to the similarities presented in Table 3.2, subsequent BLAST results suggested reductase and phosphatase proteins as the most closely related *P. vulgaris* Kelch protein candidates to KFB^T. The former analysis (Figure 3.2-2 lower) calculated these proteins to be the 47th and 137th most similar, respectively.

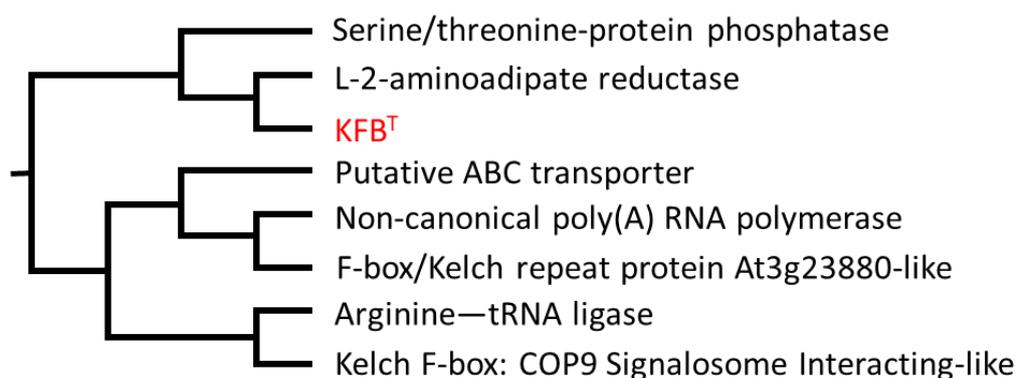


Figure 3.2-3: The KFB^T-containing clade from a phylogenetic tree of all 155 Kelch proteins identified in *P. vulgaris* with their corresponding identities.

It was noted that the 7th and 8th most similar Kelch candidates to KFB^T (Table 3.2) were situated within close proximity of each other in the *P. vulgaris* genome, separated by only one gene. Relationships between all 155 identified *P. vulgaris* Kelch genes were therefore investigated and six pairs were found to be immediately adjacent. An instance of three Kelch genes interspersed by only one other was also found (not shown).

To assess whether this was due to chance or if Kelch genes are more commonly co-localised, the GFF annotation was used to calculate average distances between all consecutive genes in the genome and compare them to the distances between these adjacent Kelch family members (Figure 3.2-4). Alignments were also carried out between these adjacent Kelch genes and the other identified Kelch candidates to investigate whether similarity was negatively correlated to the distance between them (Figure 3.2-4).

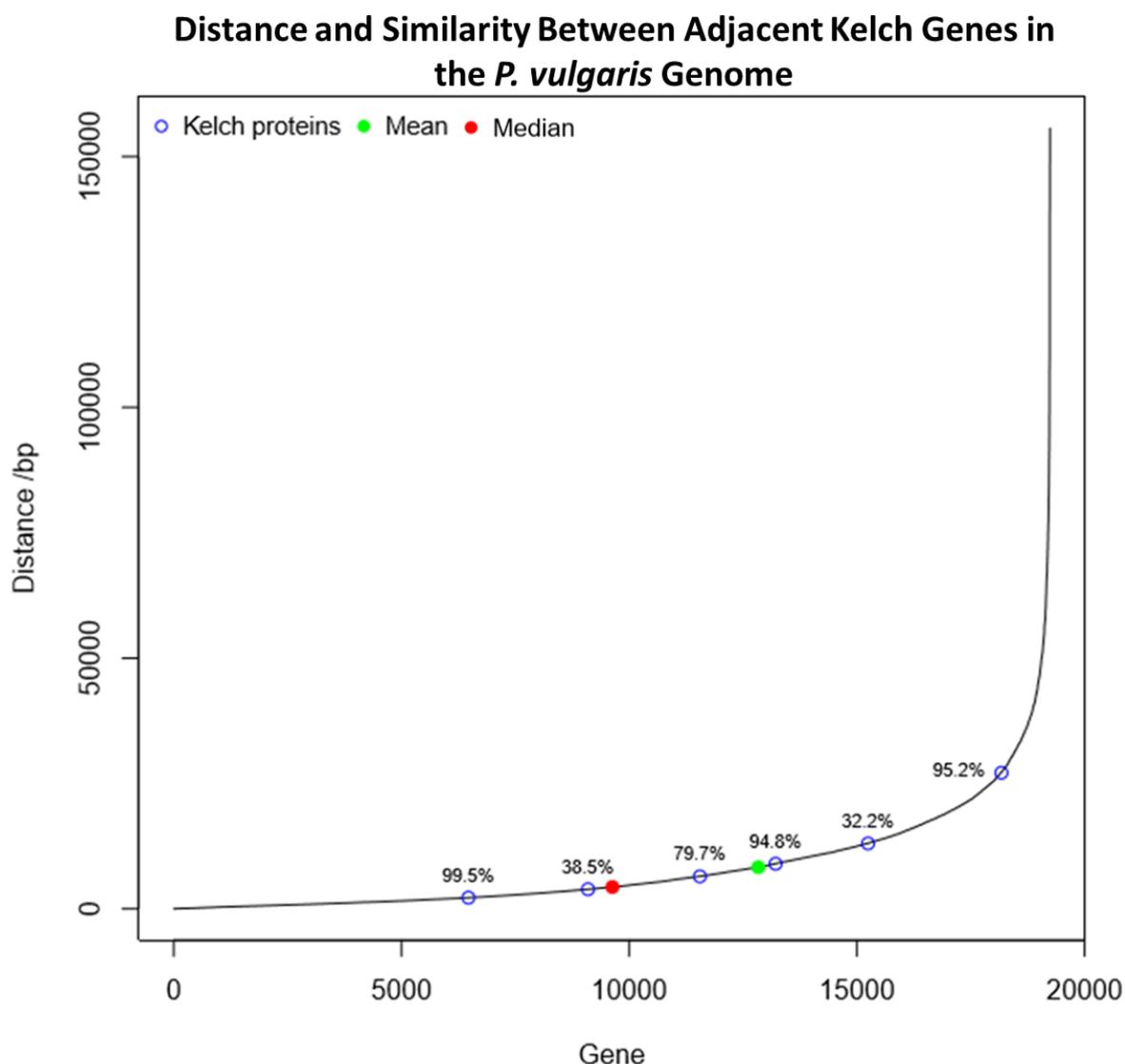


Figure 3.2-4: Similarity and distance between pairs of Kelch protein-encoding genes that are adjacent in the *P. vulgaris* genome, plotted on a line representing the distance between all adjacent genes from the genome. Adjacent Kelch genes are not significantly closer than other adjacent genes and similarity does not correlate with proximity.

Figure 3.2-4 shows that proximity between members of the Kelch family in the *P. vulgaris* genome is not significantly closer than other consecutive *P. vulgaris* genes. The figure also demonstrates that the percentage similarity between these Kelch pairs does not increase as distance between them decreases. Though some pairs were found to share as little as 30-35 % similarity, this is not a good criterion on which to measure homology here because high variance often exists between even closely related Kelch proteins (Adams *et al.*, 2000). Similarity between adjacent Kelch partners was therefore compared to their similarity against the other 154 unconnected *P. vulgaris* Kelch proteins (Figure 3.2-5).

Figure 3.2-5 demonstrates that genes encoding Kelch proteins may be less similar to their adjacent partner than they are to other Kelch family members positioned more distantly on the *P. vulgaris* genome. This further confirmed a lack of correlation between gene distance and similarity. While four of the pairs were the first or second most similar Kelch candidate to each other, other partners ranked as little as 89th or 112th in similarity (Figure 3.2-5).

Similarity Between Kelch Protein Family Members Adjacent and Non-Adjacent in the *P. vulgaris* Genome

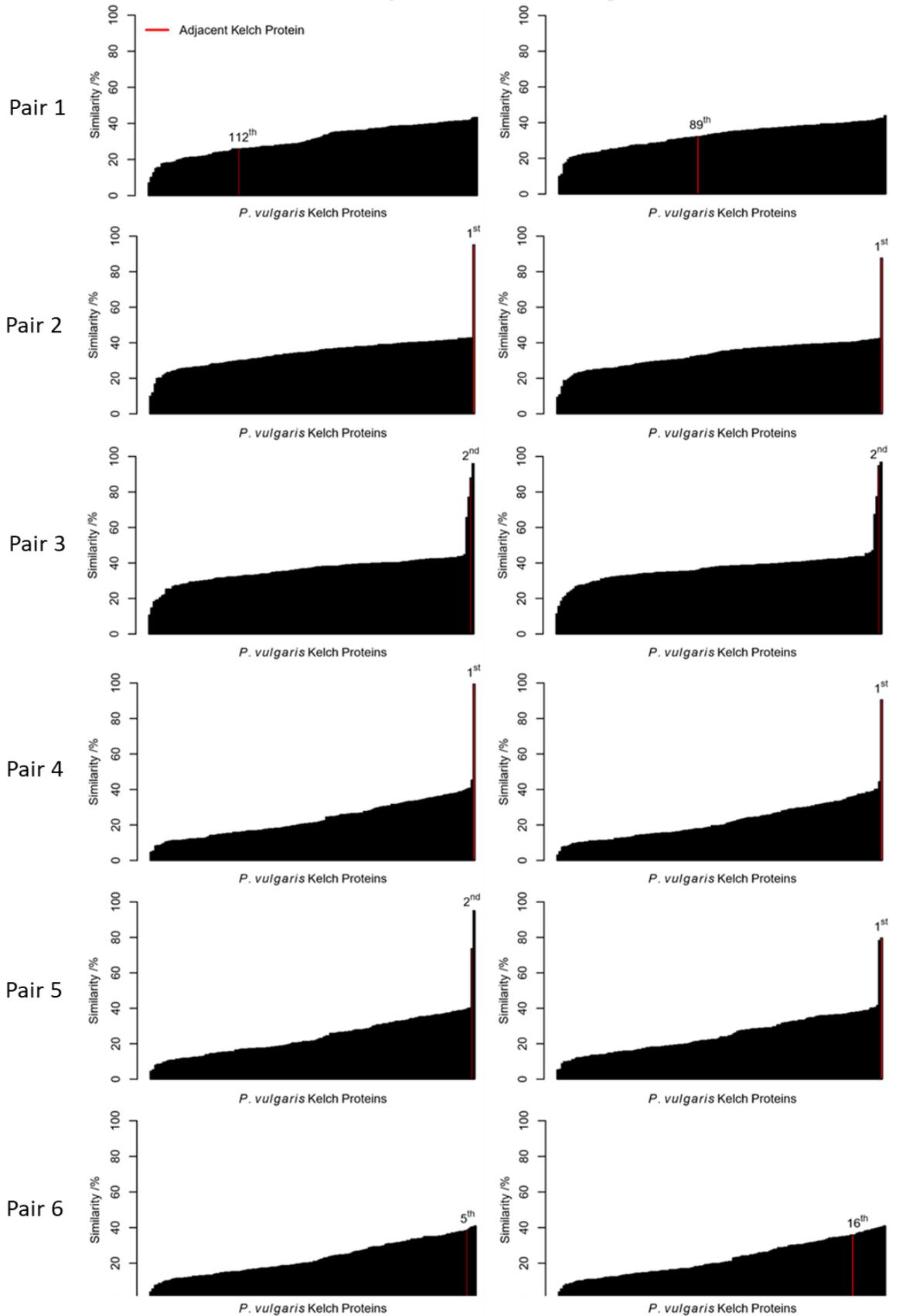


Figure 3.2-5: Six pairs of *P. vulgaris* Kelch genes adjacent in the genome and their similarity to the other 154 Kelch protein encoding-genes. The adjacent gene encoded the most similar Kelch protein 42 % of the time and was the second most similar a further 17 % of the time.

3.3 Identifying *KFB^T* in Other Members of the *Primulaceae* Family

A sequence capture experiment was used by Li & Cocker (unpublished) to identify the *S* locus genes in other species of the *Primulaceae* family. Amino acid sequences converted from contigs carrying *KFB^T* were used to assess the evolutionary selection pressure on *KFB^T* and prove its importance in heterostyly by confirming conservation across the family. Similarities along the protein sequences were compared to explore phylogeny and identify the most variable or conserved regions in each. This were used to indicate protein motifs, such as Kelch repeats or F-box domains.

3.3.1 Obtaining *KFB^T* DNA Sequences from Sixteen *Primulaceae* Species

Sequencing data from sixteen species was analysed, fifteen of which were from a sequence capture experiment by Dr Jinhong Li that utilised oligonucleotide primers designed to the 278 kb *P. vulgaris* *S* locus and its ~170 kb flanking regions to obtain the *S* loci from other *Primulaceae* species. Those species included *P. chungensis*, *P. cockburniana*, *P. concholoba*, *P. cuneifolia*, *P. halleri*, *P. incana*, *P. kewensis* (*P. floribunda* x *P. verticillate*), *P. laurentiana*, *P. prenantha*, *P. prolifera*, *P. scotica*, *P. vulgaris* (short homostyle mutant), *P. watsonii*, *P. yuparensis* and *Hottonia palustris*.

The *P. vulgaris* sample was included in the sequence capture as a positive control. Aside from the *H. palustris* thrum morph, homostyle plants were otherwise used because this experiment was originally designed to determine whether homostyle morphs in these species also arise via mutation in the *CYP^T* gene (Li *et al.*, 2016) – which has been shown to regulate style height (Huu *et al.*, 2016) – but it also generated sequences of the other *S* locus genes and provided *KFB^T* data for this investigation. The *H. palustris* material was taken from a pond near Watton in Norfolk and all other plants were sampled from Kevock Garden Plants in Midlothian, Scotland.

The contig from each sample dataset that contained the closest match to *KFB^T* was isolated by Dr Jonathan Cocker. A sixteenth *KFB^T* sequence was also included from Dr Jonathan Cocker's draft assembly of the *P. veris* thrum genome; this did not come from sequence capture but was included to maximise data quantity.

3.3.2 Similarity Between KFB^T Nucleotide Sequences from Sixteen *Primulaceae* Species

To assess whether KFB^T was present in these sixteen tested *Primulaceae* family members, percentage similarity was calculated between the isolated nucleotide sequence from each species and the *P. vulgaris* KFB^T sequence, which was here used as a template. To establish the threshold below which it would be determined that KFB^T was not present in a particular species, the *PUM^TS* locus gene was included in the analysis. This was arbitrarily selected as a negative control to demonstrate the chance similarity presented by the alignment of an unrelated gene. The script used for this analysis is included in Appendix A2.

The similarity calculations were carried out as presented in Figure 3.3.2, with use of nucleotide sequences. ClustalW (Larkin *et al.*, 2007) was used to align the KFB^T sequence from each species to that of the *P. vulgaris* template (Figure 3.3.2A). The sequence capture contigs could have contained KFB^T in either the sense or antisense direction and so alignments were performed against both the forward and reverse complement sequences. Whichever presented the greater number of matches was carried forth. The output text files were then manipulated to extract only the alignment data (Figure 3.3.2B), separate the sample and template sequences into discrete files as single lines (Figure 3.3.2C) and then re-join them as vertically aligned columns (Figure 3.3.2D) to enable inter-column comparisons and thereby allow better operation of BASH command line interface tools, which are usually written to handle files on a line-by-line basis.

Differing sizes of some query sequences caused gaps in the template sequence during alignment, to which no nucleotide aligned (Figure 3.3.2E). These gaps were discarded because the overall aim was to calculate the number of template nucleotides that appeared in the query sequence, not to count the number of query nucleotides that appeared in the template sequence. This is an important distinction that could otherwise significantly underrepresent the percentage similarity of large sequences and overrepresent it in smaller ones.

The number of matches between the query and template sequences were enumerated and converted into a percentage of 1023 (which is the nucleotide length of KFB^T in *P. vulgaris*). The percentage similarity of the KFB^T candidate from each species was presented in a bar chart (Figure 3.3.4 upper).

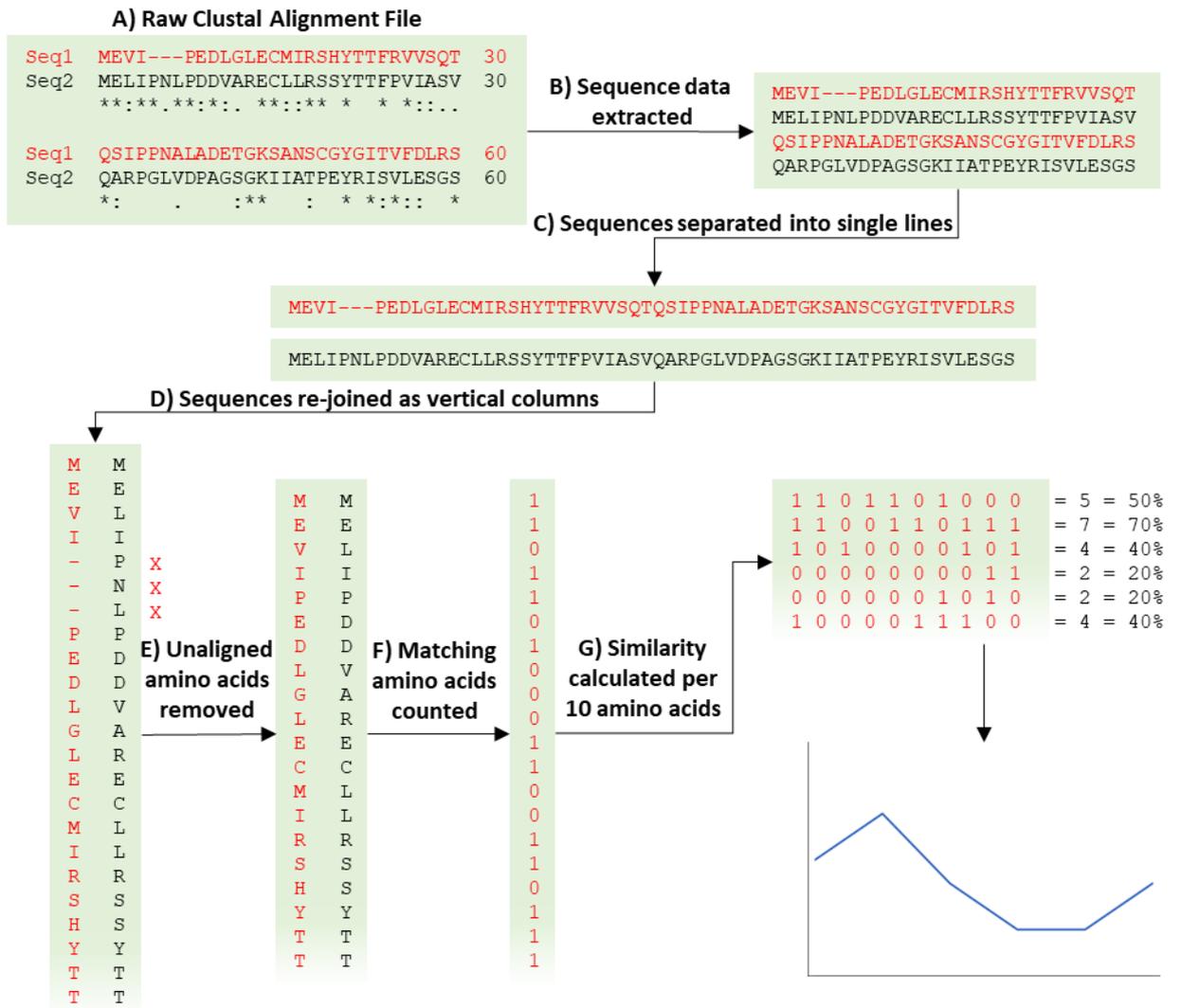


Figure 3.3.2: The process by which raw sequence alignments between KFB^T and the four AtKMD proteins were converted into graphs for comparing how similarity changes across these sequences. A) Clustal Omega (Sievers *et al.*, 2011) was used to align each AtKMD protein against KFB^T. B) The sequence data was extracted to remove labels, numbers and alignment symbols. C) The KFB^T and AtKMD sequences were separated. D) The sequences were converted to vertical columns and re-joined. E) Gaps from amino acids that did not align were removed. F) Matching amino acids were enumerated. G) Average similarity was calculated for every 10 amino acids and the data was plotted on a graph.

3.3.3 Translating *KFB^T* Nucleotide Sequences from Sixteen *Primulaceae* Species

Nucleotide alignments are affected by synonymous substitutions that do not functionally modify the primary protein structure. Analysis of the inter-species *Primulaceae* sequence capture data was therefore repeated with amino acid sequences instead. A program was written in the C programming language to translate the candidate *KFB^T* nucleotide sequences from each species. The process is outlined in Figure 3.3.3, the source code is available on GitHub (github.com/calumraine/cranslate) and the program was compiled using version 5.3.0 of GCC (gcc.gnu.org). It functions by accepting the name of an input FASTA file as an argument on the command interface. The program ignores any line that does not begin with a sequence base (in capital letters only), therefore skipping blank lines and header lines containing the ‘>’ symbol.

The program moved through the given sequence in three-base windows. Upon recognition of a start codon, a ‘print’ flag was raised and it began translating each nucleotide triplet into the corresponding single-letter amino acid code. This continued until the stop codon was found or the end of a line was encountered. A newline was triggered in the output stream to position the cursor for printing the next string.

If the end of the sequence had been reached, the process was repeated with the second and third reading frames. The input sequence was subsequently reversed and the complementary nucleotides saved in memory as a new sequence, so codons in the reverse three reading frames could be translated and queried too. If at any point the program encountered a non-base letter (such as sequencing gaps represented by ‘N’) that meant the codon could not be faithfully translated, an ‘X’ was declared instead – which could not be mistaken for any of the other single-letter amino acid codes. Where possible, the program would attempt to translate a codon if the first two bases were available.

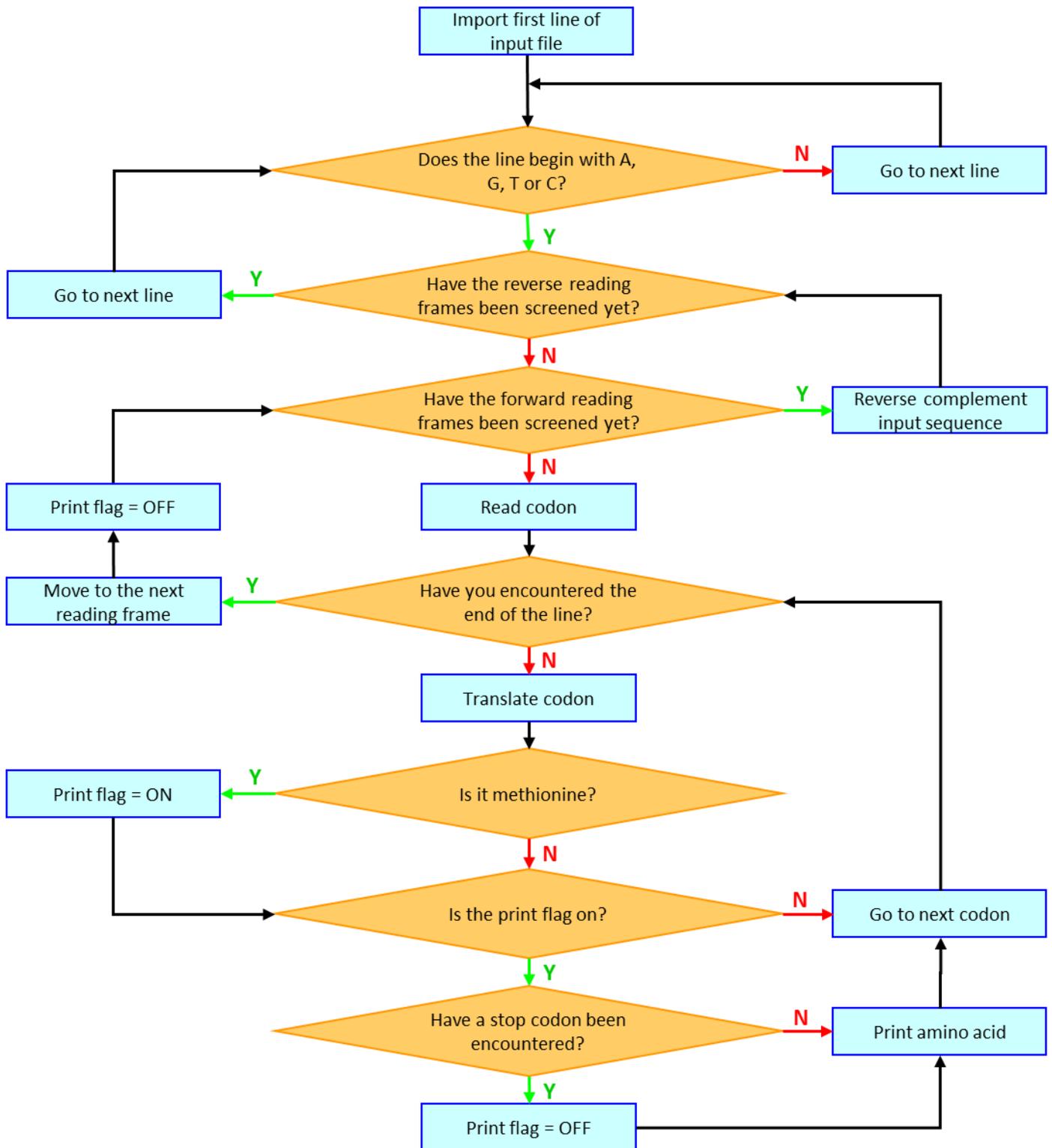


Figure 3.3.3: A program was written in the C language to find open reading frames and print their translated amino acid sequences from a nucleotide input file.

3.3.4 Similarity Between KFB^T Amino Acid Sequences from Sixteen *Primulaceae* Species

The program outlined in Chapter 3.3.3 was used to generate a file containing all the possible open reading frames from each candidate nucleotide sequence. The KFB^T protein is 340 amino acids long so the sequence closest to this length was extracted from each output file and carried forward as the final KFB^T candidate amino acid sequence for that species. The *PUM^T* coding region was translated using Clustal Omega (Sievers *et al.*, 2011) and its amino acid sequence was added to the sample set at this stage for use as a negative control. The subsequent alignment process used was the same as that in Chapter 3.3.2 (Figure 3.3.2). A final bar chart (Figure 3.3.4 lower) was produced to represent the percentage similarity between the amino acid sequences of KFB^T and its closest matches in these sixteen species from the *Primulaceae* family.

Alignments of both nucleotide and amino acid sequences suggest *KFB^T* is present in all sixteen sequence capture samples, with only minor disagreements regarding the ranking of species based on similarity (Figure 3.3.4). On average, there was only a 4.53 % reduction in similarity between alignments of the nucleotide and amino acid sequences. The most dissimilar of the sequence capture samples was *Hottonia palustris* and this was also the only sequence in which the start codon did not align to that in *P. vulgaris*. The next start codon was 22 amino acids further upstream in the *H. palustris* sequence.

Similarity Between KFB^T From Various *Primulaceae* Species

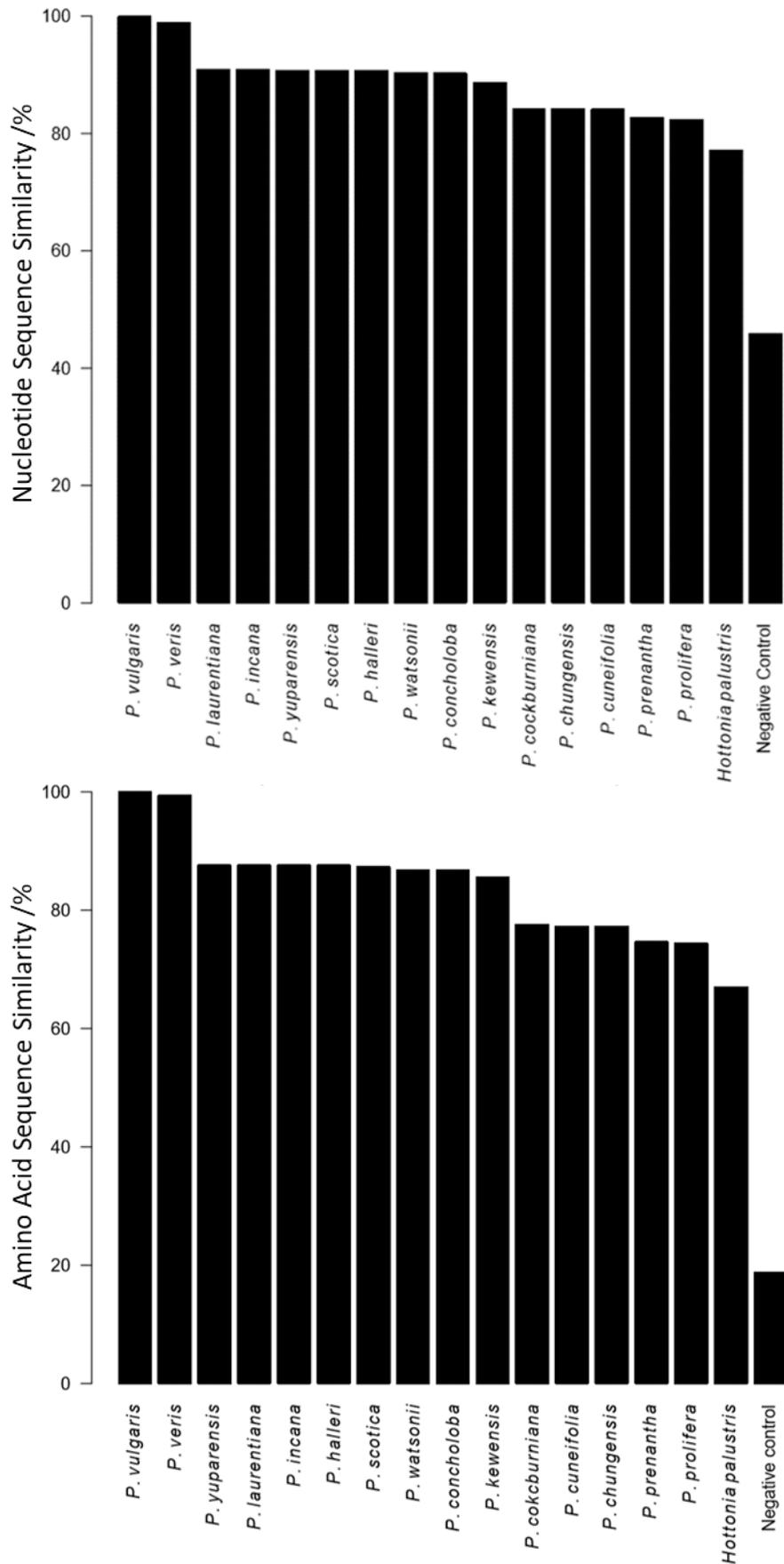


Figure 3.3.4: Identification of KFB^T in sixteen species from the *Primulaceae* family. Presence was confirmed in all. Similarity of nucleotide (upper) and amino acid (lower) sequences differed by <5 %, on average. The negative control gene is PUM^T from the *P. vulgaris* *S* locus, here used as an unrelated gene to demonstrate chance similarity.

3.3.5 Conserved Regions in KFB^T from Sixteen *Primulaceae* Species

The ratio of synonymous to non-synonymous mutations is an indicator of evolutionary selection for purification or diversification (Zhang, 2006). A consensus sequence was generated by aligning the sixteen *Primulaceae* KFB^T nucleotide sequences to each other (Appendix A3). Each species was subsequently aligned to the consensus sequence and a script was written to detect synonymous and non-synonymous substitutions within each codon (Appendix A4). The overall ratio for each species was presented on a stacked bar chart (Figure 3.3.5-1) and the average mutations across all sixteen nucleotide sequences were also plotted (Figure 3.3.5-3).

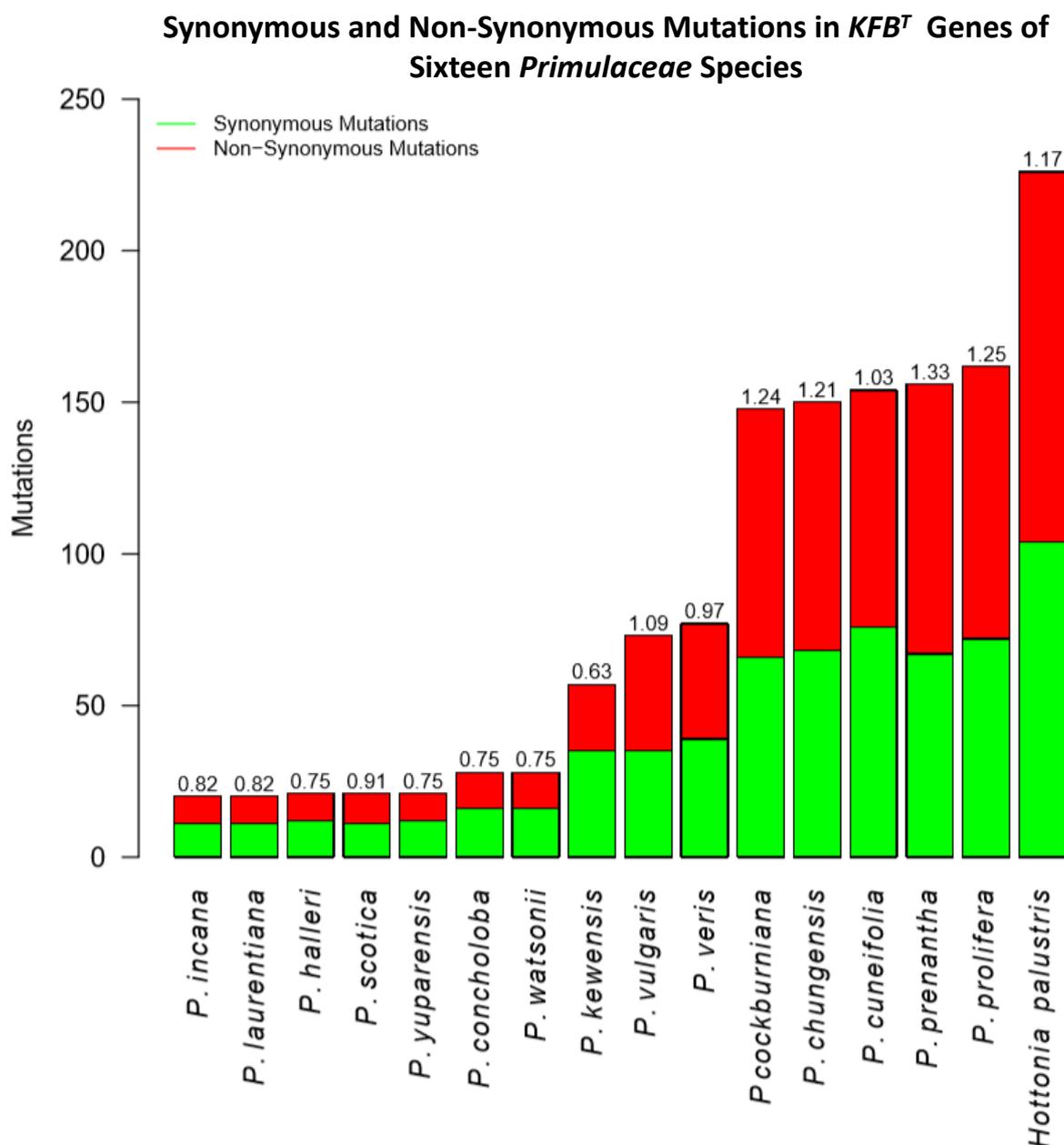


Figure 3.3.5-1: Ratio of non-synonymous to synonymous substitutions between KFB^T nucleotide sequences from sixteen *Primulaceae* species. The *H. palustris* sequence contained the most mutations but *P. prolifera* and *P. cockburniana* demonstrated the highest diversifying selection.

The lowest mutation counts (Figure 3.3.5-1) were observed in eight species that appear to share the most distant common ancestors from the clade containing *P. vulgaris*, according to the phylogenetic tree in Figure 3.3.5-2. These samples also demonstrate purifying selection whereas *P. chungensis*, *P. cockburniana*, *P. prolifera* and *P. prenantha* demonstrated diversification with ratios between 1.21 and 1.33 for non-synonymous to synonymous mutations.

Clustal Omega (Sievers *et al.*, 2011) was also used to produce a phylogenetic tree of the *Primulaceae* species from their KFB^T amino acid sequences (Figure 3.3.5-2). The ranks are taken from the similarity assessment between the KFB^T amino acid sequences in Figure 3.3.4 (lower). The Clustal results are largely coherent with those similarity ranks except *P. vulgaris* and *P. veris* appear to share a more recent common ancestor with the most dissimilar species than the former alignments would suggest. Clustal also placed *P. kewensis* closer to *P. vulgaris* than the Figure 3.3.4 analysis proposed.

Clustal Phylogenetic Tree	Similarity Rank
<i>H. palustris</i>	16
<i>P. cuneifolia</i>	12
<i>P. chungensis</i>	13
<i>P. cockburniana</i>	11
<i>P. prenantha</i>	14
<i>P. prolifera</i>	15
<i>P. veris</i>	2
<i>P. vulgaris</i> (Short homostyle)	1
<i>P. vulgaris</i> (wildtype)	NA
<i>P. kewensis</i>	10
<i>P. concholoba</i>	9
<i>P. watsonii</i>	8
<i>P. scotica</i>	7
<i>P. halleri</i>	6
<i>P. incana</i>	5
<i>P. laurentiana</i>	4
<i>P. yuparensis</i>	3

Figure 3.3.5-2: Clustal Omega (Sievers *et al.*, 2011) was used to align KFB^T amino acid sequences from 16 *Primulaceae* species and produce a phylogenetic tree. The rank value (right) is based on previous similarity calculations and here allows comparison between the two methods.

Contigs containing *KFB^T* from the sequence capture experiment had an average length of ~1.6 kb. However the *P. vulgaris* short homostyle contig was almost 48 kb long and so was screened for other genes. It was found to contain ten potential open reading frames that were longer than 100 amino acids. The BLASTp results indicated that these did not represent novel identified genes but likely retrotransposon insertions or genes bordering (*Cfb^{TL}*) or internal (*PUM^T* and *CYP^T*) to the *S* locus (Table 3.3.5).

Table 3.3.5: A sequence capture experiment to target the *S* locus was carried out in 15 *Primulaceae* species and contigs containing *KFB^T* were identified. The *P. vulgaris* short homostyle contig was approximately 30,000 times larger than any others. It was found to contain ten potential open reading frames longer than 100 amino acids.

Open Reading Frame	Length /aa	BLAST Identity
1	509	<i>PUM^T</i> (<i>P. vulgaris</i>)
2	398	Gag-Pol polyprotein (<i>Trifolium pratense</i>)
3	340	<i>KFB^T</i> (<i>P. vulgaris</i>)
4	312	Glycosyl transferase (<i>Pseunocardiales bacterium</i>)
5	220	Transposon Tv3-I Gag-Pol polyprotein (<i>Vitis vinifera</i>)
6	145	No matches
7	130	Hypothetical protein (<i>Thiohalocapsa</i> sp. ML1)
8	128	<i>CYP^T</i> (<i>P. vulgaris</i>)
9	116	Hypothetical protein (<i>Auxenochlorella protothecoides</i>)
10	107	<i>Cfb^{TL}</i> (<i>P. vulgaris</i>)

Figure 3.3.5-3 presents the average number of both mutation types per ten codons along the KFB^T sequences. The graph demonstrates a clear crossover of synonymous mutations above non-synonymous mutations across the central gene regions; an area directly coherent with the middle three repeats of the previously identified Kelch domain (Figure 3.1). This represents evolutionary effort to conserve the integrity of this domain, unlike its diversified bordering regions that present far higher ratios of non-synonymous to synonymous mutations.

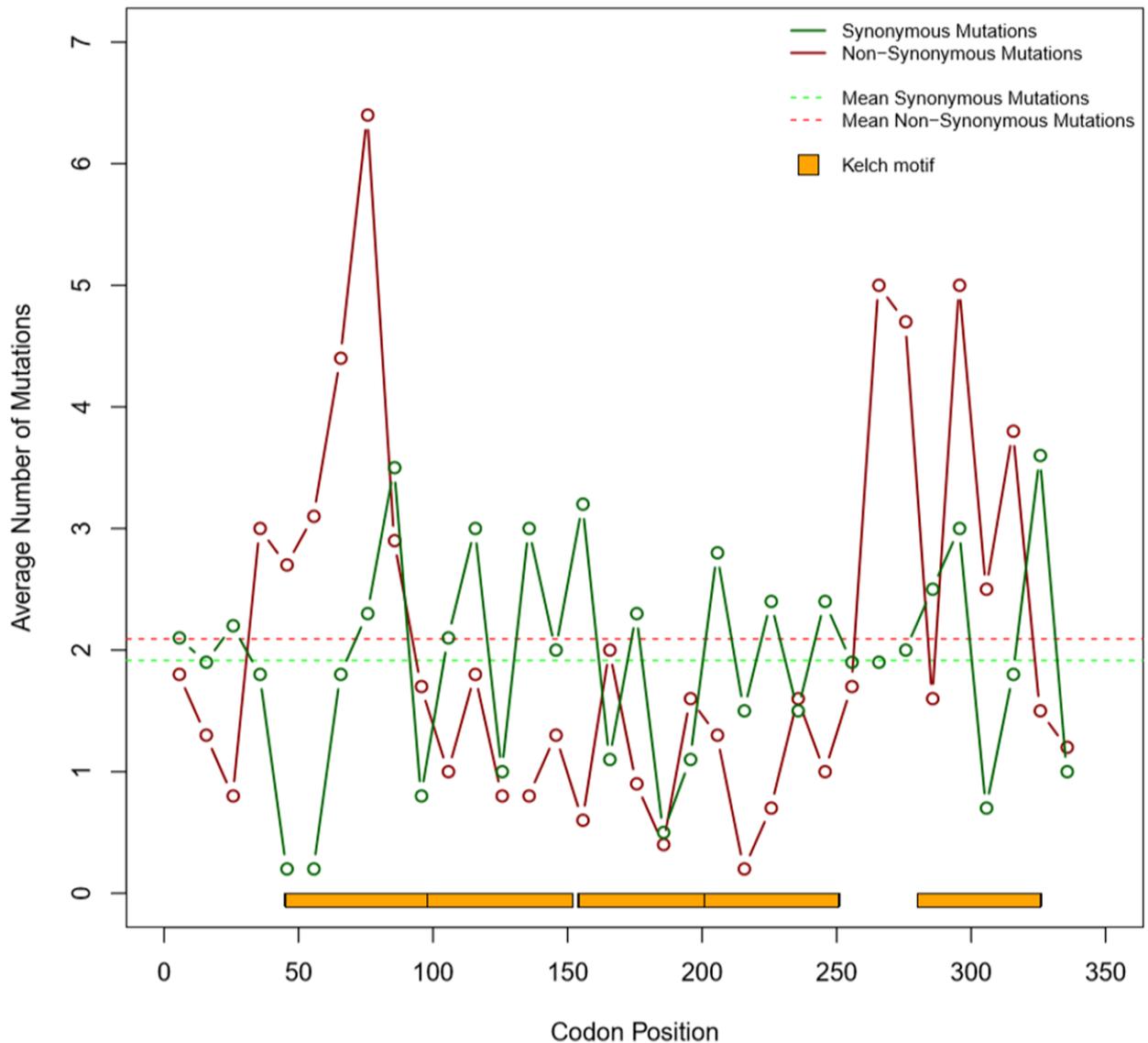


Figure 3.3.5-3: Average number of synonymous and non-synonymous mutations across KFB^T sequences from sixteen *Primulaceae* species. A central region of purifying selection corresponding to the middle three Kelch repeats was identified.

The large red peak immediately preceding the Kelch domain in Figure 3.3.5-3 corresponds to the space between the fourth and fifth Kelch repeat, in which a fractional Kelch motif was identified (Figure 3.1). The 5'-end of the gene near the start codon shows more modest counts of non-synonymous mutations and this is exaggerated further where the 3'-terminus exhibits resistance to change. Sequence corresponding to the second Kelch blade (codon 98 to 152) presented strongest purifying selection. The lowest number of both mutation types was observed in the third motif, in which the ratio between them suggests the presence of neutral evolutionary selection pressure.

3.4 Homology of KFB^T to *Arabidopsis* KMD Proteins

Similarity between the four *Arabidopsis* KMD (AtKMD) proteins and *P. vulgaris* KFB^T was calculated across their amino acid sequences. This allowed identification of the most similar AtKMD protein to KFB^T and also highlighted the more variable or highly conserved regions of their amino acid sequences. Identification of the closest known homologue to KFB^T could provide early clues regarding its function and help direct the lab studies. Figure 3.3.2 outlines the process used to make these calculations and produce a graph. The BASH script written to carry this out is included in Appendix A2.

The KFB^T coding region was obtained from the *P. vulgaris* genome (Li *et al.*, 2016) and the nucleotide sequence was converted into an amino acid sequence via the ExPasy online translation tool (Gasteiger *et al.*, 2003). Amino acid sequences of the four *Arabidopsis* KMD proteins were taken from the TAIR online database (www.arabidopsis.org). Each AtKMD was aligned individually to the KFB^T amino acid sequence via the Clustal Omega online multiple sequence alignment tool (Sievers *et al.*, 2011).

Table 3.4: Amino acid sequence similarity between KFB^T and the four *Arabidopsis* KMD homologues. Similarity is presented as the number of KFB^T amino acids that align to KMD and these matches are also provided as a percentage of the KMD amino acid sequence length, shown in column 2.

Protein	Coding Sequence Length /aa	Similarity to KFB ^T Amino Acid Sequence	
		Matches	%
KFB ^T	340	NA	NA
KMD1	354	113	31.92
KMD2	359	120	33.42
KMD3	409	110	26.89
KMD4	418	108	25.84

The KFB^T amino acid sequence was found to convey most similarity to KMD1 and KMD2 than KMD3 or KMD4 (Table 3.4). Although KMD3 and KMD4 are the longest of the four homologues, they have the fewest number of matches and therefore the lowest overall percentage similarities. The function of KFB^T was presently unknown but these four homologues had been studied in *Arabidopsis*. Literature regarding the most related candidate was therefore reviewed to infer functional predictions for KFB^T.

3.5 Identification of Conserved KFB^T Domains via Comparative Genomics

To highlight conserved domains within KFB^T, three sets of alignments were carried out between KFB^T against the four *Arabidopsis* KMD homologues (Chapter 3.4), the other 154 Kelch proteins identified in *P. vulgaris* (Chapter 3.2) and the KFB^T paralogues identified in fifteen other *Primulaceae* species (Chapter 3.3). Similarity along the KFB^T amino acid sequence helped to identify potential functional domains alongside regions of variability in the protein.

In accordance with Figure 3.1, Clustal (Larkin *et al.*, 2007) was used to align KFB^T against the query amino acid sequences. The raw alignment output was copied into a text file (Figure 3.1A) and the sequence data was extracted (Figure 3.1B) to remove any surplus lines, labels, numbers and alignment symbols. Tools commonly used in the Linux command line interface are usually designed to handle files line-by-line, so scripts are far smoother when they make comparisons between individual columns instead of between separate lines. The sequence for each sample was therefore extracted into a single line (Figure 3.1C) before both being re-joined as vertical columns (Figure 3.1D). Lines containing amino acids that did not align to KFB^T were removed (Figure 3.1E). The number of matches at each position were then enumerated (Figure 3.1F) and the average similarity across every ten amino acids was calculated (Figure 3.1G) to generate a smoother graph line when displaying the results (Figure 3.5).

Amino Acid Sequence Alignments Between KFB^T and Various Homologues

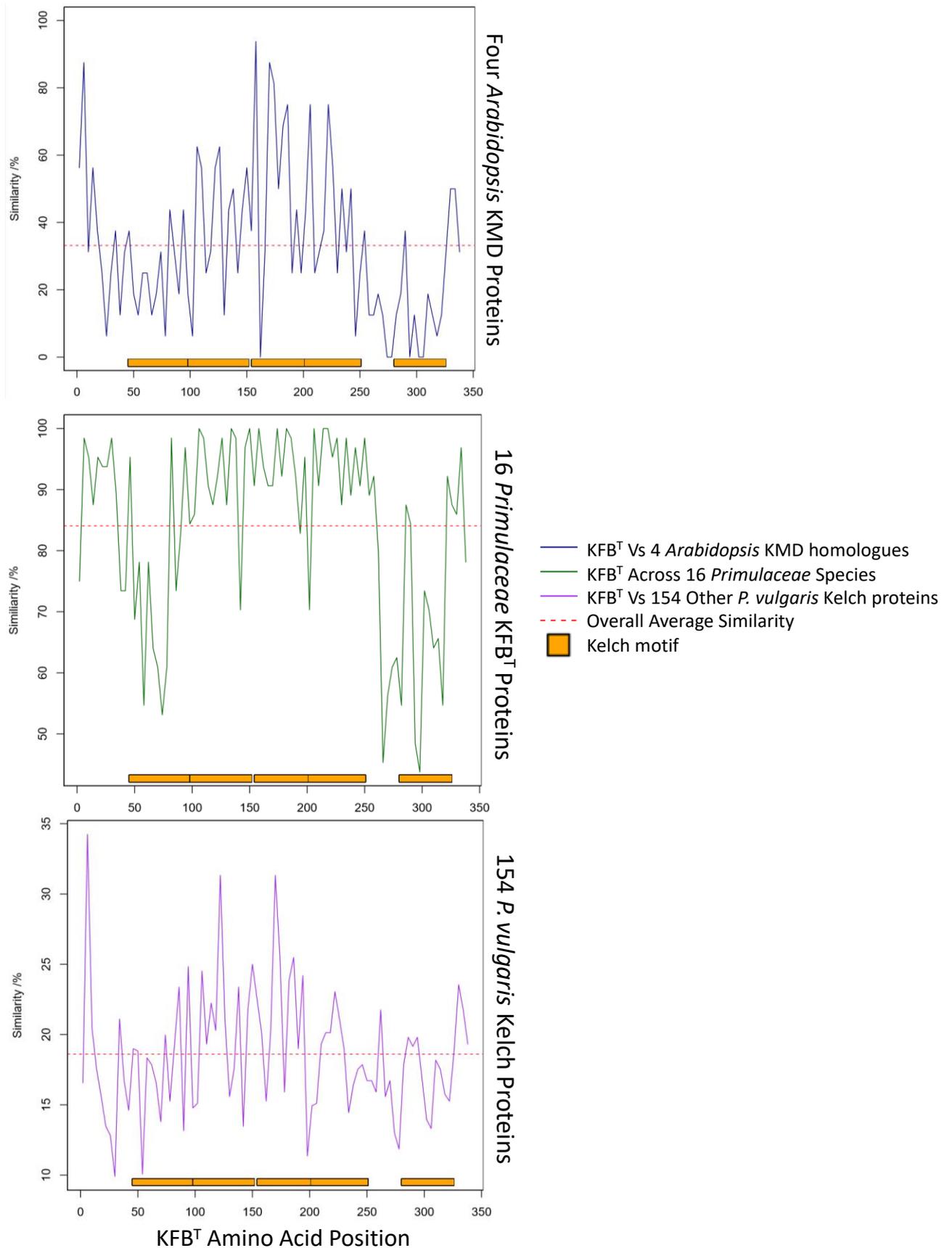


Figure 3.5: Amino acid sequence alignments were carried out between KFB^T and: Upper) four *Arabidopsis* KMD homologues, Centre) KFB^T from across sixteen *Primulaceae* species, Lower) 154 other Kelch proteins identified in *P. vulgaris*. They all demonstrate a central region of raised similarity across the Kelch repeats.

A central region of raised similarity was common to all three alignments (Figure 3.5), which corresponds to the middle three blades of the KFB^T Kelch domain (Figure 3.1). Similarity reduced over the fourth motif when compared across the entire *P. vulgaris* Kelch protein family but remained above average. Although a spike of similarity was observed in the fifth Kelch repeat, the region between the fourth and fifth motifs (found to contain a partial Kelch motif) was highly variable in all three alignments. Similarity across the first proposed Kelch repeat in KFB^T was not strong across any of the datasets and this was also the weakest match to the Kelch consensus sequence (Figure 3.1). Similarity consistently increased at the C-terminus of the protein but an even greater increase was observed at the N-terminus.

3.6 Transcriptional and Translational Regulation of *KFB^T*

Upstream and downstream regions of the *KFB^T* gene were screened for potentially significant regulatory elements, such as secondary structures, open reading frames, a transcription start site and promoter components. Annotating binding sites and modelling secondary structures can provide understanding of the pathways that a gene may be involved in or respond to. Such insights aid in better hypothesis formulation to direct experimental design. Placement of the transcription start site was also required for deciding how much of the upstream sequence to include in a vector designed to contain the *KFB^T* promoter region (Chapters 4.8 & 5.3).

3.6.1 Manual Screen for Promoter Elements and Transcription Start Site

The *KFB^T* transcript was isolated from a thrum transcriptome assembled by Dr Jonathan Cocker (*et al.*, 2018) and the sequencing read was found to end 1275 bp upstream of the start codon, which highlighted that the 5'-UTR could be at least 252 bp longer than the 1023 bp *KFB^T* coding region itself. Following this, manual analysis of the upstream region was carried out.

Transcription start sites are usually overlaid by an initiator element positioned 25 to 30 bp downstream of a TATA box (Juven-Gershon *et al.*, 2010), often located in a GC-rich region (Lewin, 2001). For identification of the *KFB^T* transcription start site, the 1900 bp region upstream of the start codon was queried for a potential initiator element linked to a TATA box by this distance. An initiator element consensus sequence was used for the query: [C/T][C/T]AN[T/A][C/T][C/T] (Javahery, 1994). Although identified in mammalian genomes, its functionality has been confirmed in plants (Nakamura, 2002).

A total of 184 TATA box candidates and 43 potential initiator elements were identified in the 1900 bp upstream of the *KFB^T* start codon. The TATA boxes were identified as mixed strings of eight adenine or tyrosine bases but only five of them were between guanine or cytosine residues, suggesting the others were simply strings of AT repeats and not true TATA boxes. All candidate initiator elements obeyed the consensus sequence but only one (Figure 3.6.1-1C) was paired with a TATA box (Figure 3.6.1-1B) positioned 29 bp upstream within a GC-rich region. This potentially placed the *KFB^T* transcription start site at the adenine base underlined (Figure 3.6.1-1C) 1644 bp upstream of the start codon.

The region was also screened for other common promoter elements (Porto *et al.*, 2014), such as Downstream Promoter Elements (DPE): [A/G]G[A/T][C/T][G/A/C] (Burke & Kadonaga, 1996). A total of 35 potential DPE fragments were identified (not shown) and three were 24, 30 and 37 bp downstream of the identified initiator element (Figure 3.6.1-1C). True DPE sequences are approximately 30 bp downstream of the TSS (Burke, 1996). No attempts to identify suitable upstream CAAT box (Bucher, 1990), Sp1-binding GC box (GGGCGG), B recognition element ([G/C][G/C][G/A]CGC) or gibberellin response site (AAACAGA) were successful (Porto *et al.*, 2014).

It is possible that not every gene in the *S* locus elicits direct function in floral heteromorphy but instead co-ordinates other members of the *S* locus. As *GLO^T* is a known transcription factor with a MADS-box domain, the upstream region of *KFB^T* was queried for a CA₆G box (CC[A/T]₆GG), which is the consensus sequence for MADS-box binding sites (Reichmann *et al.*, 1996). A match to this consensus was found 1877 bp upstream of the *KFB^T* start codon (Figure 3.6.1-1A).

A) **CArG** TCTCTAGAGACCACACAAATACCA **CCATATAAGG**ATTTTATT -1860
 CTCCAACATACTTTTGTTTAATTCATCTAGCAAAAAAGTTTCATCCACATCCAAAAGGCT -1800
 TAGATTTGCGATGTTTTGCATAGATTGGTGATGATTGGTGTGCGATTAAGGACCTAATTG -1740

B) **TATA** TTGATCATGTGGTTTGGGTGGATGACGGATGCCGATGACTGTTCACTGCGGTGGGC **TAT** -1680

C) **Initiator** **ATAAT**GGTGCTCCGACGTTCTTTGGTCACTTGT **TCAGATT**AGGGATTCTTTACTTACTCG -1620
 GATCTGGTTCAAGGATATTGGAGGTATGAAAGTCCAGTAAGTGTGGGACAAAGTTGAG -1560
 ATGGCGTTCTGGAGGAGGAGAGAATATGAAAAAAAAAATGATTTATATATTATTATTTTC -1500
 CTAACCTATAGATGTACTTATCTAAATGATGTGCACATATTGGTAGATACACGTATAAC -1440
 TCTCTTATCTAGCTGCATTTGTCTTATATTTCGTCTTATATCGAACTATTAGTTGGGAATT -1380
 TGGGATAATATAAGTAAATCTTTACAAATCTAATGAATTGAAACCCTAAACCCTAAATTA -1320

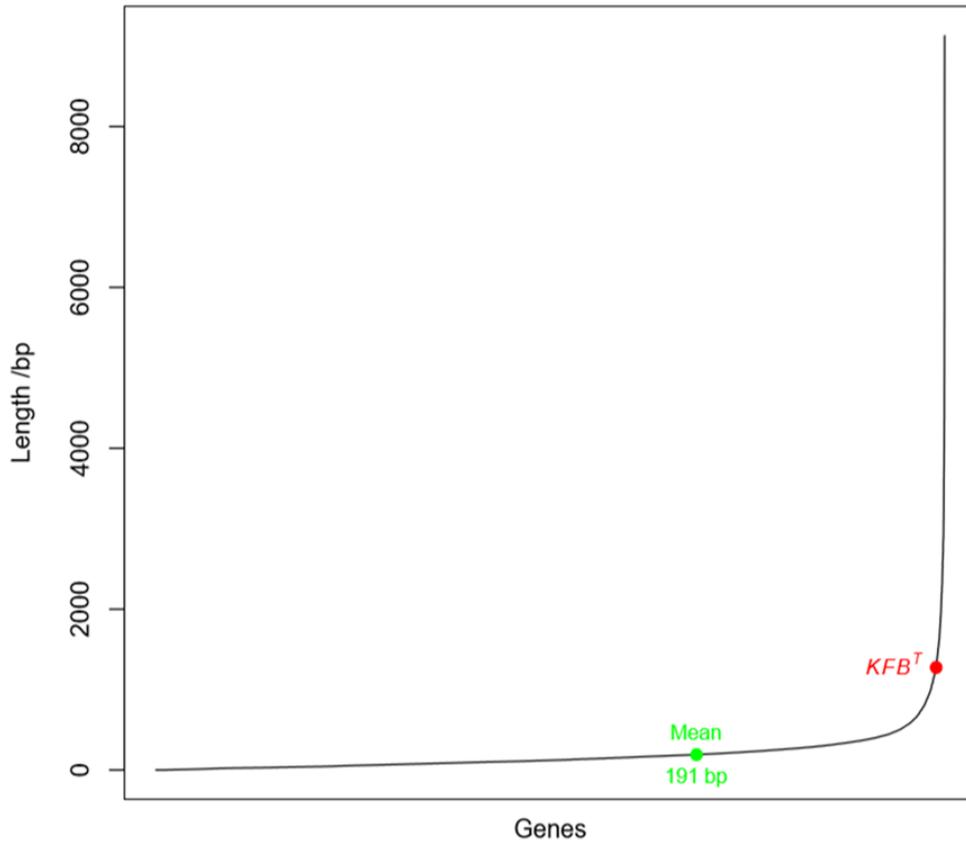
D) **Transcriptome** CTAACAACCTCAATTTTAATAAATAAAAAATTTATTTCTTTA **G**GTATGATTAATATATA -1260
data end GAATAGAGTAAAAACTGTTAATATCGTCTAAATAAATAAATTGAATCTAATAAACTCA -1200
 ACTTCGTATGGGTTAGAATCTAATTAGAATTTTATTTTATTTTAGTATAAAACTAATTAC -1140
 TTGAAAAGATTCTTAAAGTATTTAATAAGTTTAATAATGAATTTACATTAAAAAACGAT -1080
 TTAACCTATATTTATTAATTTGTAGCAAGTTGTTTTGTAGTTTATATCATTTTTAAATT -1020
 TGCTTTTATAAAATGTTTTTTTTTTCAGTTACACATAATTTTATAAACTTTCAATATTTT -960
 TATATAAGTGTGTAATTTATAAAATTTCAAACGTTTATTGTAAAGTGTAAACTATACA -900
 TAATAAATGGAACTTTAAAATGGGTTCTATTCAAAAAATGTAACGATGGATCACTAAAA -840
 CAATCCAAATTTAAAATTTAGACCAAAATCAGTAGCAAAATATGATTAACGTTACAAATAT -780
 TTATAGTCACCCCGTTCCAAATAGAGTTATTCTAATTTGGGATGTATCTGGAAATGAAAC -720
 TTTTCCGAAATTGATGTCTAATACATATGCATTTTTACCCTATTACTCTACCTGTGTACA -660
 TAGTAAAAGAAATATTATTTAACTAAATGAAAGGATTTTTTAACAAATAAATATAGAA -600
 TCACTAAATTTTCATGAATAATTGGGTTCTCCAAGCGTCTAATGAATATAACGAAGTAG -540
 TACAAAACAAATGTGCGGTGGACACTTTAATATATTTCTCCAGTACTCGATCAACACAGTT -480
 TTTAGAAGTATAAGATTAGATTAATAAATTTTGATTAACAGATAGTCCCACATGCATA -420
 ACTGCCCAAAGGCACCGCGGGATAAAGAGAGCGAGAGAGAACTTTGCCACAATTTAC -360
 CATCATCTATTCCTTCGGCAAAAACCTTGCTATCTGAATGTACACTAATCTTTTTTCGAA -300
 TTCTTCAAGTCTTTTAATGACACACTAGTTGAAAGAACCTAGCTAGTAGCAGGGTCTTAC -240
 TCCCGCTACAGTTACCAAGAGTTACTTCTTCATCTTGACAGTTGACACGTAATAAATTA -180
 ATATTAAATCTTTTCTTTTTTCTGTGTTATATATATTTTCAGCATATTATAAGTACCTCA -120
 CTTACTGGTCTCGGTATTTTCATATCGTCATATCCCTCCCACTTCGTGAATAAAACCTCA -60

E) **Start codon** TTGACAAAACCTTAAATAAATATCTAACTATTTTCATAGCCAATCTTTGAAAATAAT **ATG** 0

Figure 3.6.1-1: The upstream region of *KFB^T* was searched for common promoter elements. A) CArG consensus sequence for MADS-box transcription factor binding sites. B) Potential TATA box, 30bp upstream of an initiator element. C) Possible initiator element, the third base of which could be the *KFB^T* transcription start site. D) 5'-end of the *KFB^T* sequencing read from the *P. vulgaris* transcriptome. E) *KFB^T* start codon.

This candidate transcription start site (Figure 3.6.1-1C) is a further 367 bp from the end of the *KFB^T* sequencing read in the *P. vulgaris* transcriptome; a total of 1644 bp upstream of the start codon. To assess the hypothesis that this was significantly long, the *KFB^T* coding sequence and 5'-UTR lengths were compared to those of all other *P. vulgaris* genes (Figure 3.6.1-2). This put the proposed *KFB^T* 5'-UTR within the longest 1.15 % of all *P. vulgaris* genes. In addition, length of the *KFB^T* coding sequence was almost average (Figure 3.6.1-2 lower).

5'-UTR Lengths of *P. vulgaris* Genes



Coding Sequence Length of *P. vulgaris* Genes

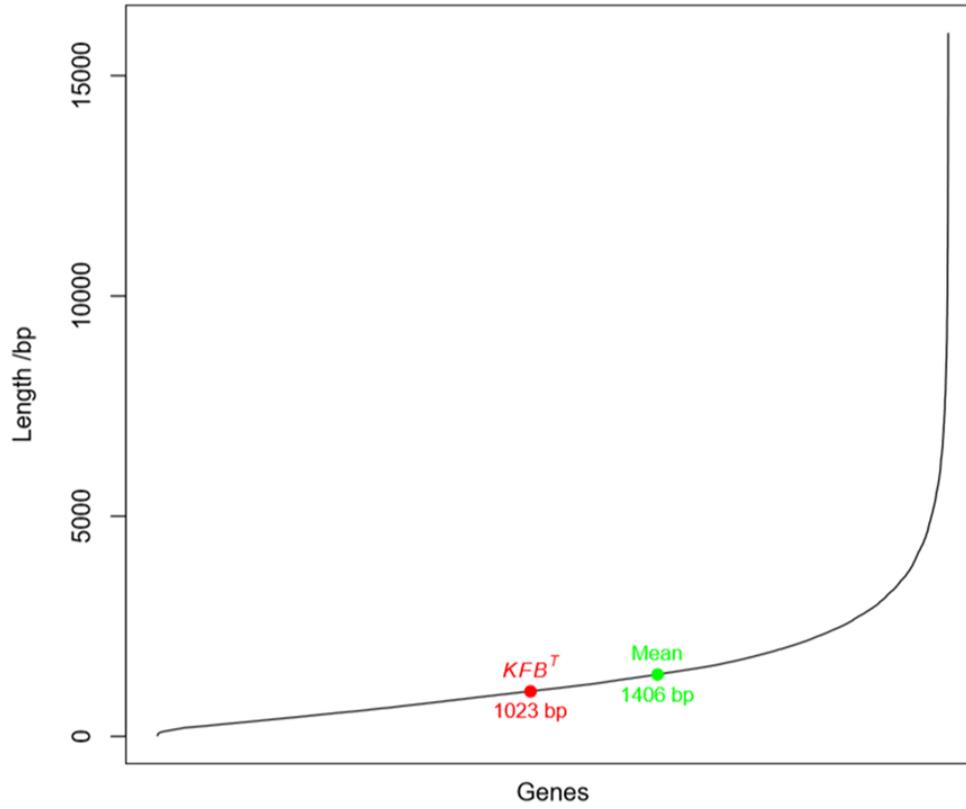


Figure 3.6.1-2: Lengths of the *KFB^T* 5'-UTR (upper) and coding sequence (lower) were compared to those from all other *P. vulgaris* genes.

3.6.2 Self-Complementarity within *KFB^T*

It was hypothesised that such a long 5'-UTR of *KFB^T* (Figure 3.6.1-2 upper) could exhibit a regulatory role. To investigate this further, a dot matrix alignment was carried out between the *KFB^T* nucleotide sequence and its own reverse complement to identify self-complementary regions that may have regulatory function. The screened area included untranslated regions surrounding the *KFB^T* coding sequence. This *KFB^T* sequence was broken into consecutive kmers and each one was used to scan the reverse complement sequence for matches. The dot matrix presented the frequency and locations at which each kmer occurred. Dots that appear close to the x=y line represent matches that are near to each other in the sequence, while dots further from this line represent greater distance between the two matching regions.

Any matching kmer between these two sequences could form a potentially self-complementary region *in vivo*. The process was repeated using kmer lengths from 6 up to 40 bases and the longest match was selected for each position and presented on the final dot matrix (Figure 3.6.2-1). Aside from identifying self-complementary regions, areas of high repetition presented themselves as clusters of dots on the graph. The script used to generate this data is included in Appendix A5.

A heat map of GC content was generated for the samples by using a sliding window that analysed each sequence in consecutive 100 bp kmers to enumerate the frequency of guanine and cytosine occurrences. This allowed the corresponding GC content to be identified for any part of the sequence, meaning areas of interest on the graph could quickly be identified as either AT repeats or something more complex.

The *KFB^T* coding sequences corresponded to an area of increased GC content (Figure 3.6.2-1). Self-complementary matches were less dense within this region than outside of it (Figure 3.6.2-1). Aside from a self-complementary 13 bp fragment between 169 bp into the *KFB^T* coding region and 451 bp upstream of the start codon, no other matches longer than 9 bp were identified within the coding sequence.

There were two significant matches between the *KFB^T* 3'-UTR and 5'-UTR; a 15 bp match linking position +1384 with -1359 and a 14 bp match linking position +1865 with -957. A palindromic 14 bp fragment was identified 20 bp downstream of the *KFB^T* stop codon: ACATTTTAAAATGT. A second perfectly palindromic fragment of the same length was also identified in the *KFB^T* 3'-UTR, 444 bp downstream of the stop codon: TCAGTTATAACTGA.

Transcription start sites are commonly found in areas of high GC content (Lewin, 2001). One predicted transcription start site of *KFB^T* is 1644 bp upstream of its start codon (Figure 3.6.1-1C) and is positioned in a GC-rich region according to the heat map in Figure 3.6.2-1, providing further evidence that this could be the true transcription start site of *KFB^T*. However, multiple GC-rich regions also occur in the 1 kb immediately preceding the *KFB^T* start codon, which is where a transcription start site would usually be expected to be found. Two main clusters of repetitive regions appeared 0.5-1.5 kb and 2-3 kb upstream of the *KFB^T* start codon. These largely corresponded to areas of low GC content in the heat map and are therefore indicative of AT repeats in the promoter.

Dot Matrix Alignment Between KFB^T Nucleotide Sequence and Its Reverse Complement

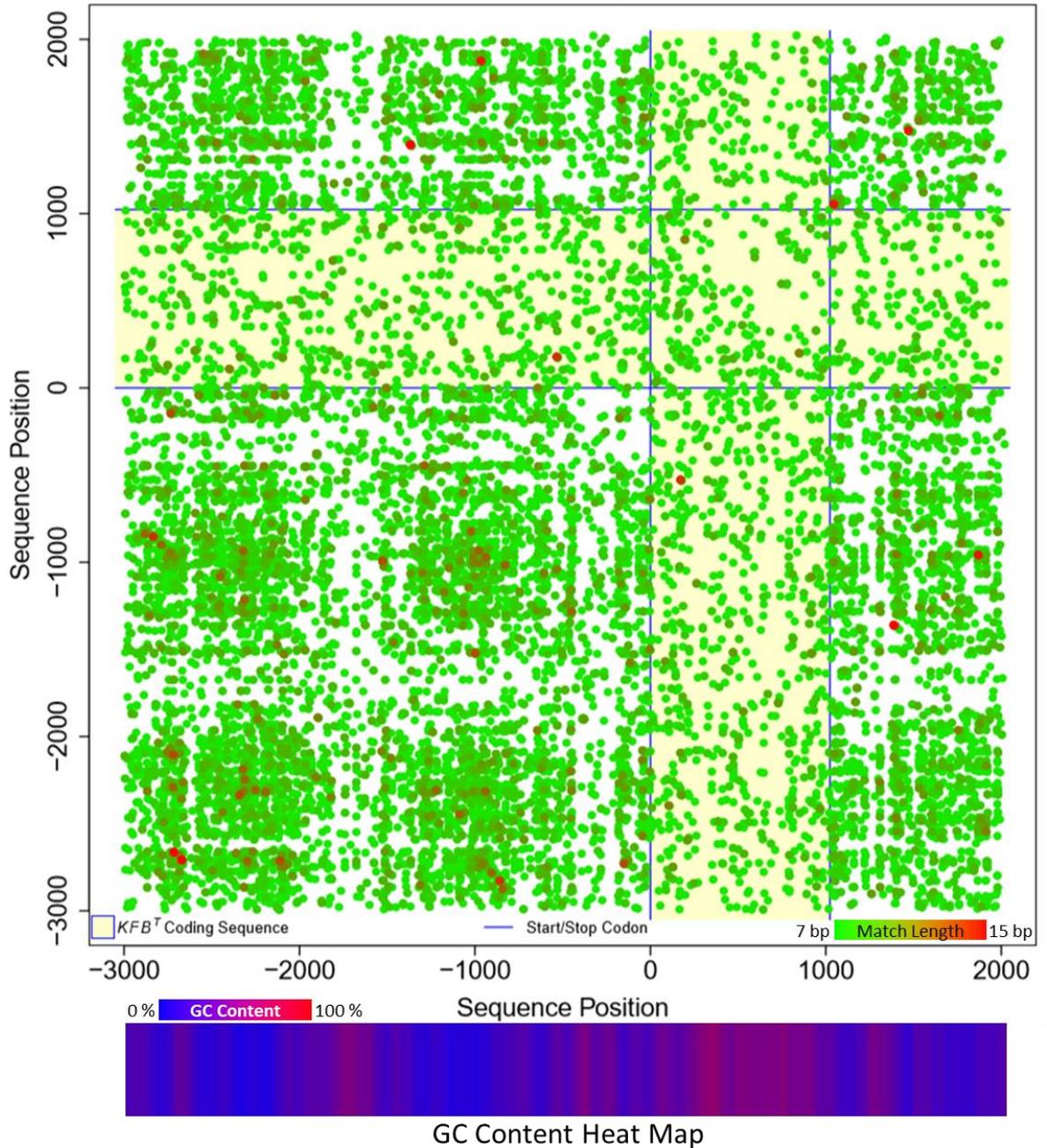


Figure 3.6.2-1: A dot matrix alignment between the KFB^T nucleotide sequence and its reverse complement to identify self-complementary matches and repetitive regions (upper). Matching sequences between 7 bp and 15 bp in length are displayed. Co-ordinates on both axes refer to 5'-3' positions on the forward sequence. A GC content heat map corresponding to the nucleotide sequence was also generated (lower).

The graph allowed identification of self-complementary regions across a range of distances, however dots closest to $x=y$ are nearer in the sequence and present a higher chance of forming RNA hairpins. The sequence was subsequently re-screened for potential stem-loop structures (Appendix A6). Stem lengths of 5 to 15 bp were used based on the *KFB^T* dot matrix alignment results. Loop lengths between 3 and 20 nucleotides were screened because those under 20 nucleotides are most common (Schudoma *et al.*, 2010). Upstream and downstream regions of *KFB^T* were also queried for alternate open reading frames in the 5'-3' orientation.

The results were mapped over the *KFB^T* nucleotide sequence alongside results of the RNA hairpin analysis (Figure 3.6.2-3). Figure 3.6.2-2 depicts the distribution of stem and loop lengths for each identified RNA hairpin. Overall, 232 fragments mapping across ~71 positions were identified (Figure 3.6.2-3) with the potential to form RNA hairpins and 98.77% of those had a stem length of 5 or 6 nucleotides (Figure 3.6.2-2). Figure 3.6.2-2 depicts the distribution of stem and loop lengths for each identified RNA hairpin.

Stem and Loop Lengths of Identified RNA Hairpins Surrounding *KFB^T*

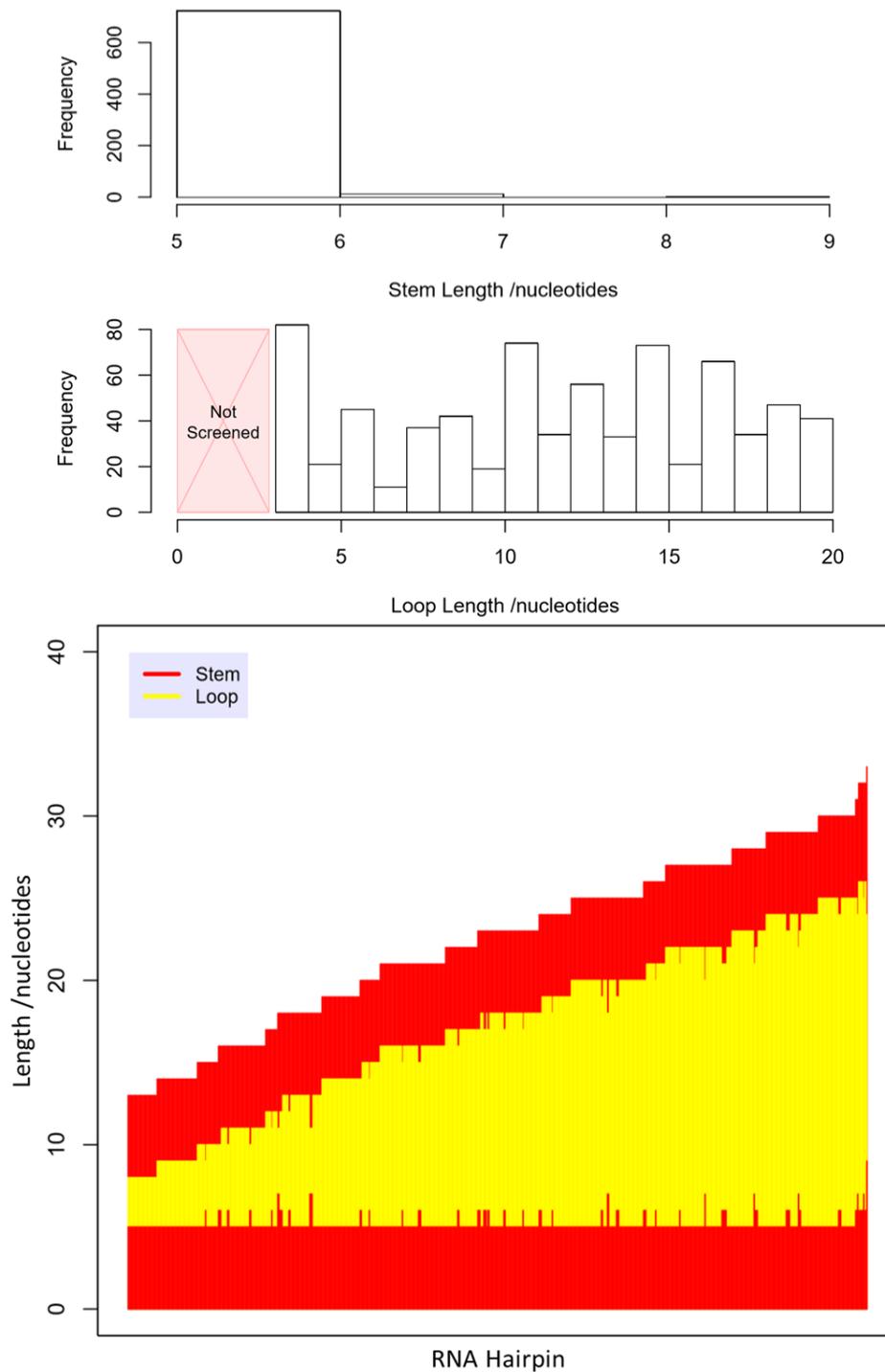


Figure 3.6.2-2: Distribution of stem and loop lengths from potential RNA hairpin structures identified within and around the *KFB^T* nucleotide sequence. The query searched 3000 bp upstream of the *KFB^T* start codon and 200 bp downstream of its stop codon. Stem fragments between 5 and 15 nucleotides were screened. Loops between 3 and 20 nucleotides were included.

One potential RNA hairpin, with a 5 bp stem and a loop length of 8 nucleotides, was found to span the *KFB^T* start codon (Figure 3.6.2-3). A further three were located in the *KFB^T* 3'-UTR. They also had 5 bp stems and possessed loops of 19, 15 and 14 nucleotides each. There were no hairpins identified around the transcription start site predicted in Chapter 3.6.1 but fourteen were found within the region of a second potential start site (discussed below; Figure 3.6.3-2).

A total of thirty open reading frames at least 10 amino acids in length in the 5'-3' orientation were identified in the untranslated areas surrounding *KFB^T*. Those in excess of 20 amino acids were mapped to the *KFB^T* nucleotide sequence alongside the 232 predicted RNA hairpin structures (Figure 3.6.2-3). The longest consisted of 64 amino acids and was situated ~200 bp upstream of the true *KFB^T* start codon.

RNA Hairpin Structures and Open Reading Frames Surrounding KFB^T

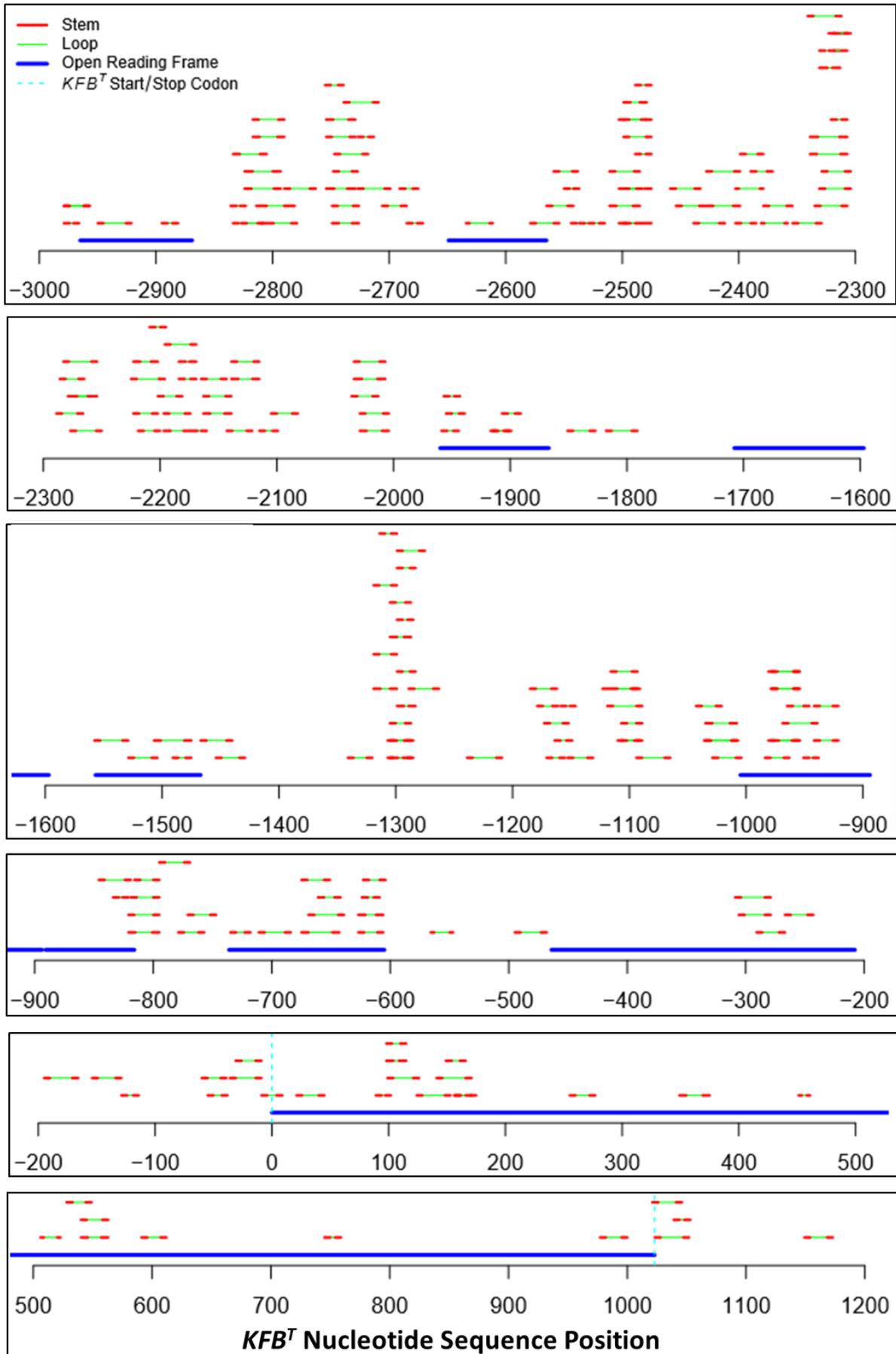


Figure 3.6.2-3: Potential RNA hairpin structures identified in and around KFB^T . Stem lengths (red) between 5 and 15 nucleotides with linker regions (green) between 3 and 20 nucleotides were assessed. Alternative open reading frames in the 5'-3' direction are also presented (blue).

3.6.3 RNA-Seq Screen for Transcription Start Site

RNA-Seq reads from pin and thrum with both long and short homostyle samples had been aligned against the *P. vulgaris* genome to allow quantification of KFB^T expression levels in various floral morphs (Chapter 3.7). A secondary benefit to this alignment was the potential discovery of previously unidentified transcripts in or around KFB^T . The StringTie (version 1.3.3; Pertea *et al.*, 2016) assembly and quantification tool was used on the Sequence Alignment Maps from HISAT (from Chapter 3.7.1). The output files contained a table of exons and transcripts predicted from the RNA-Seq data of each sample. All transcripts predicted for the scaffold on which KFB^T is known to be situated were extracted. An R package (version 5.3.0; www.r-project.org) was used to plot positions of these transcripts from each sample onto a graph. Start and end locations on the scaffold for each read were adjusted in relation to KFB^T – with positive and negative values indicating upstream or downstream positions. These were used as the x-coordinates for each transcript and a common y-value was assigned to those from the same sample set (Figure 3.6.3-1).

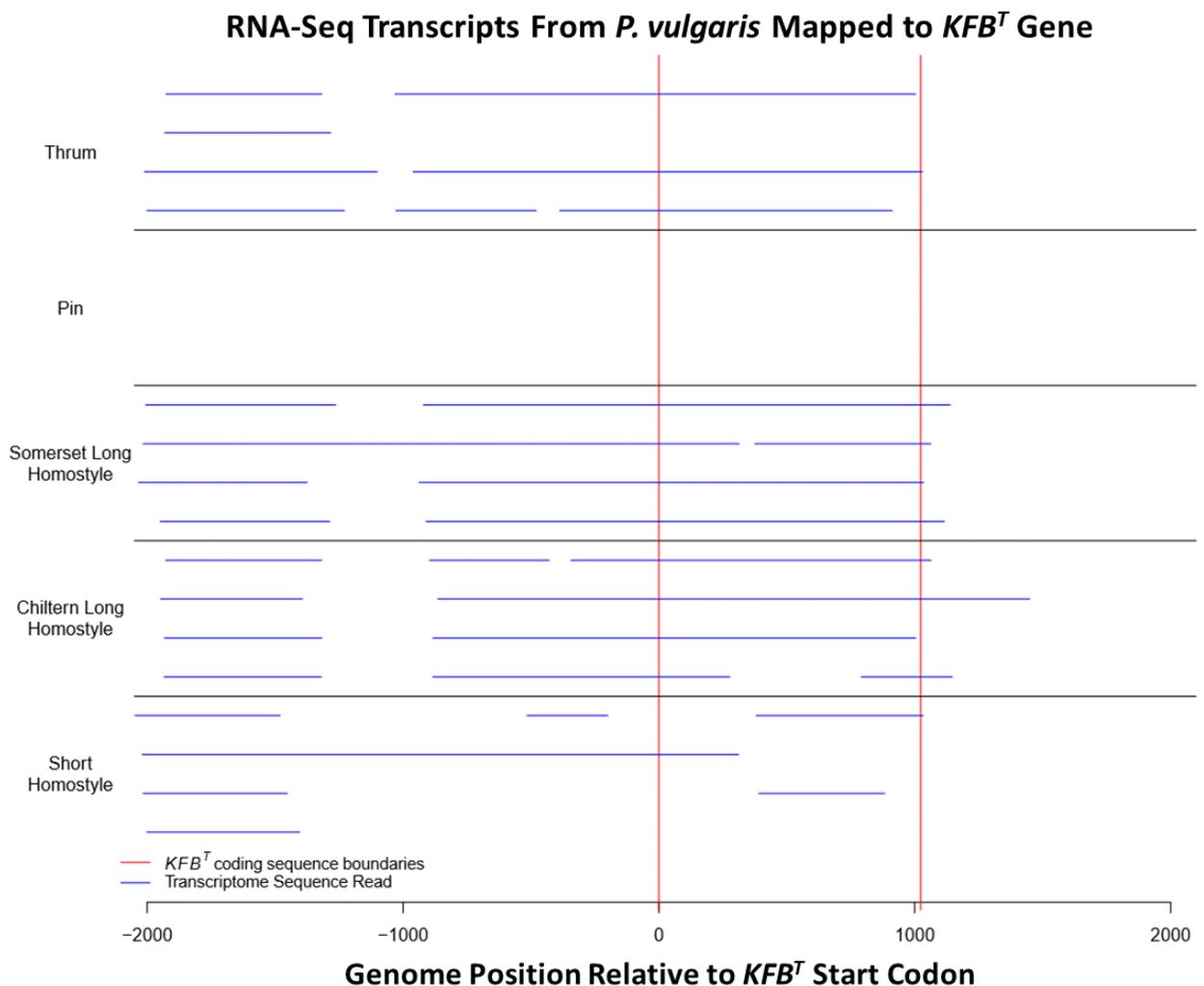


Figure 3.6.3-1: Position and length of RNA-Seq reads that mapped to the *P. vulgaris* KFB^T gene.

Assembled reads in the majority of samples appear to continue ~1 kb upstream of the start codon with a second ~500 bp transcript after an intervening ~350 bp gap. However, one short homostyle read and another from the Somerset long homostyle sample both span the entire 2 kb distance into the *KFB^T* coding region. This is even further than the potential transcription start site predicted (Chapter 3.6.1) 1644 bp upstream of the *KFB^T* start codon (Figure 3.6.1-1C).

Figure 3.6.3-2 displays transcript positions that mapped to the wider *S* locus scaffold and not only the *KFB^T* gene. As *KFB^T* is at the end of the *S* locus, downstream transcripts were outside of this unit and are therefore common to all samples whereas upstream reads did not appear in any of the pin samples. This supports the current model that the *S* locus is not present in the pin genome (Li *et al.*, 2016).

RNA-Seq reads also mapped to the fellow *S* locus genes, *CYP^T* and *PUM^T* (Figure 3.6.3-2). Reads upstream of *CYP^T* were 200-1456 bp in length and all only mapped to various bacterial artificial chromosomes made from the *P. vulgaris* genome for an alternative experiment by Dr Jinhong Li. These were discarded as insignificant and one was a clear spurious misalignment that had mapped to the pin sample. The remaining two *S* locus genes are further upstream on a different scaffold and are therefore not shown.

RNA-Seq Transcripts Mapped to *Primula vulgaris* Genome

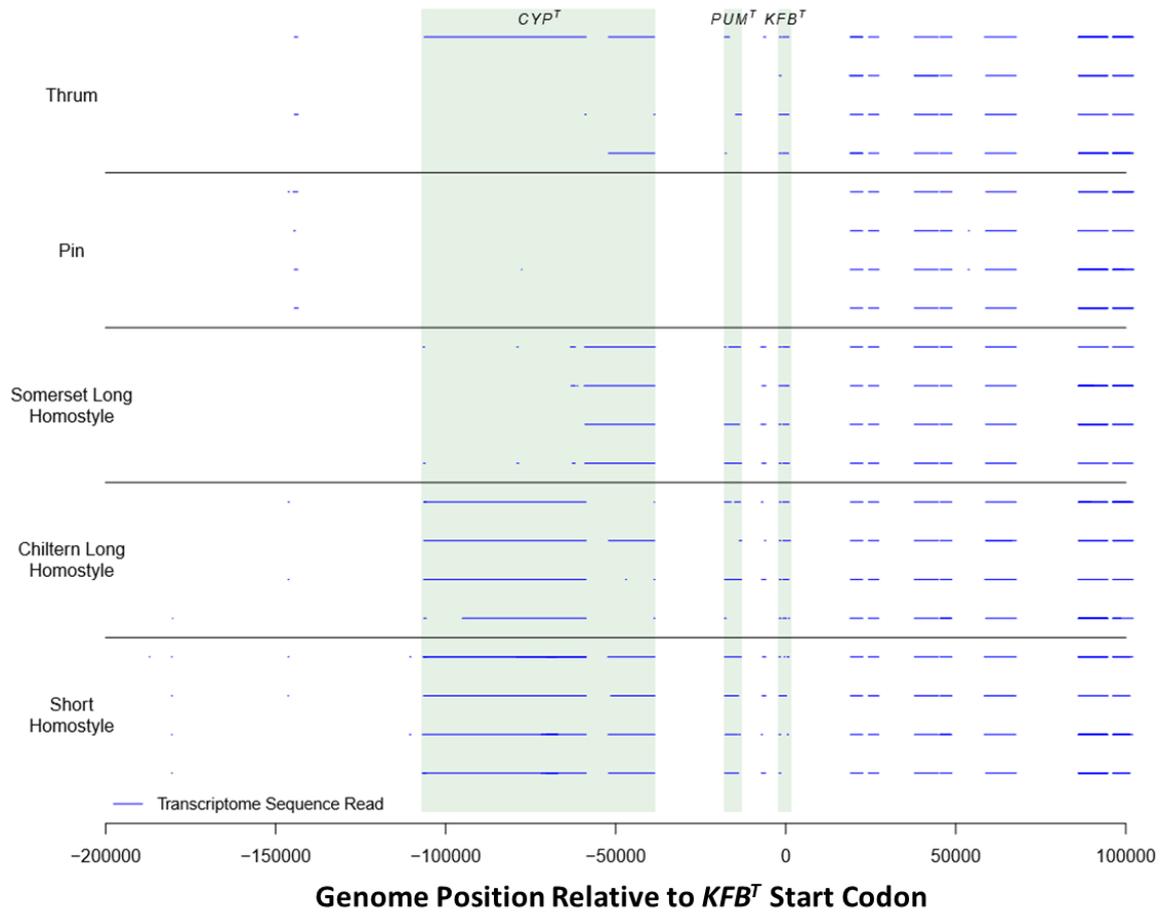


Figure 3.6.3-2: Position and length of transcripts assembled from RNA-Seq that mapped to the *P. vulgaris* genome scaffold containing *KFB^T*. Two other *S* locus genes were identified alongside *KFB^T*. Reads to the right of *KFB^T* are outside of the *S* locus.

Although output files from StringTie had been used to annotate transcript locations in and around *KFB^T*, nothing was known about their coverage. This offered a third way to visualise expression and predict the transcription start site of *KFB^T*. The HISAT package (Kim *et al.*, 2015) had been used to map RNA-Seq reads to the *P. vulgaris* genome (Chapter 3.7) and this had produced sequence alignment maps. The depth files associated with these sequence alignment maps were analysed to infer transcript expression levels from their coverage values. Mean coverage was calculated between the four biological repeats of thrum, short homostyle and two long homostyle varieties (Figure 3.6.3-3). Transcripts situated within 2000 bp on either side of the *KFB^T* start codon were included.

The upstream transcript in Figure 3.6.3-3 – approximately 2 kb away from the *KFB^T* start codon – demonstrated significant expression, higher than the *KFB^T* coding region itself in all but the Somerset long homostyle sample. A peak containing the *KFB^T* open reading

frame commenced approximately 900 bp upstream of the start codon (Figure 3.6.3-3), which is a more expected position for the transcription start site.

Around 300 bp after the start codon, depth momentarily dropped to zero in the homostyle samples before resuming. Depth values in Figure 3.6.3-3 are a mean average across four biological repeats for each sample but the individual datasets were each checked for this premature transcription termination. It was confirmed that all eight biological repeats from Chiltern and Somerset long homostyle morphs exhibited this termination ~300 bp into the *KFB^T* coding region.

The four short homostyle biological repeats also demonstrated this loss of coverage; one of which terminated even earlier at position ~200 and another presented a complete loss of expression across almost the entire first half of the coding region. Only one of the thrum replicates appeared to lose coverage at this position, with another showing no reduction in depth at all and two more demonstrating approximately 50 % loss. Furthermore, the homostyle datasets included a technical RNA-Seq repeat (thrum did not) and therefore had greater chance of containing a read to span this gap but this did not occur.

Some transcripts also ceased 200 to 500 bp upstream of the *KFB^T* start codon. This is most notable in the short homostyle sample presented in Figure 3.6.3-3 but was also observed in two thrum samples and one Chiltern long homostyle. This did not occur in the Somerset mutant line. Coverage declined rapidly in all flower morphs after the stop codon at downstream position 1023 (Figure 3.6.3-3). The coverage data in Figure 3.6.3-3 suggests a transcription start site approximately 900 bp upstream of the *KFB^T* start codon. This is a far more reasonable proposition than the 1644 bp 3'-untranslated region previously proposed (Figure 3.6.1). According to the RNA-Seq coverage data (Figure 3.6.3-3), the transcription start site from the previous prediction (Chapter 3.6.1) is positioned in the middle of what appears to be a short ~500 bp transcribed region. This expression could explain why the TATA box was surrounded by a GC-rich region (Figure 3.6.1).

RNA-Seq Coverage of KFB^T

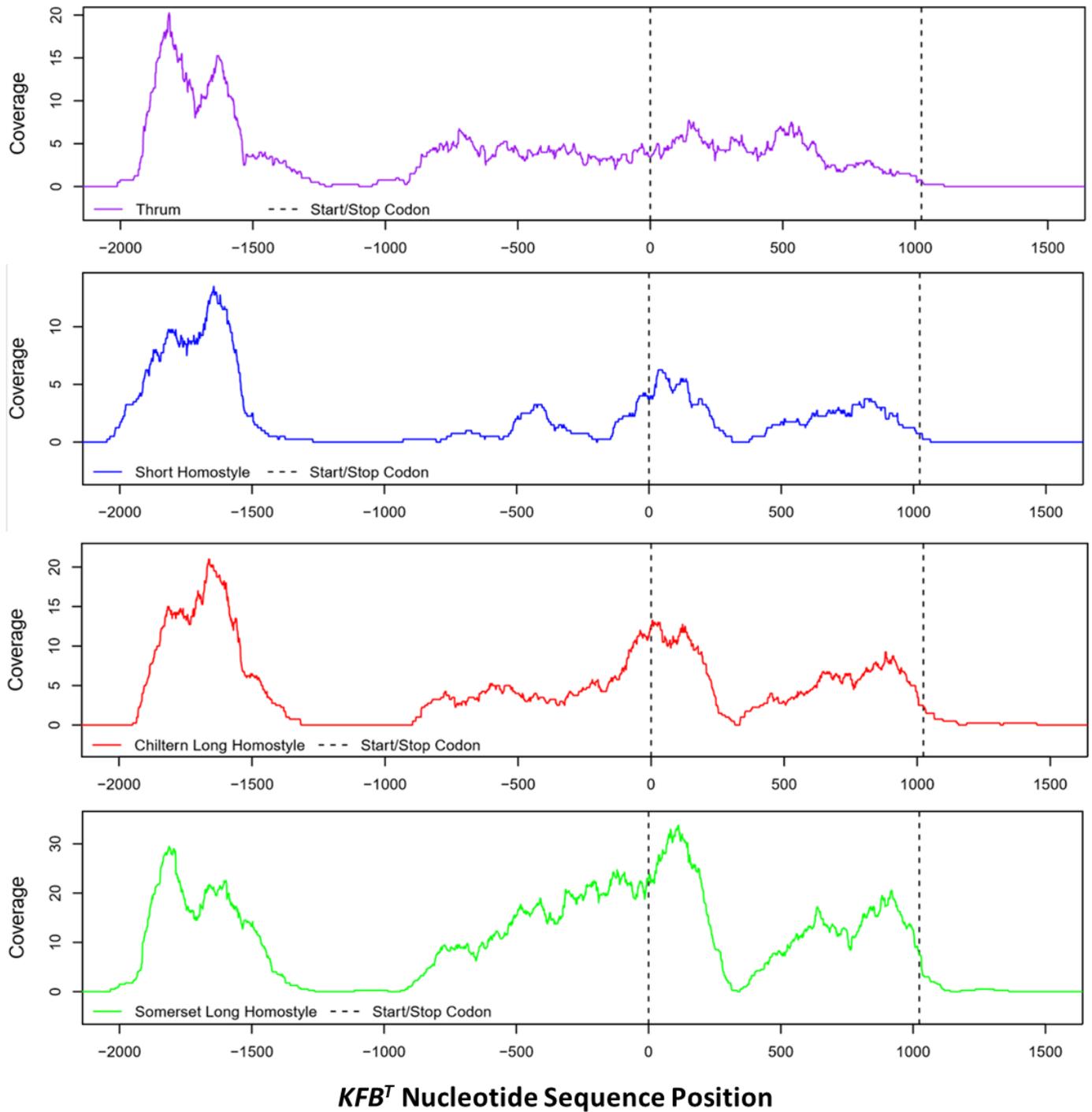


Figure 3.6.3-3: Mean coverage of RNA-Seq reads across the KFB^T region in four floral morphs of *P. vulgaris*. A near total loss of transcript depth in homostyle samples was identified in the first quarter of KFB^T .

3.7 Differential Expression of *KFB^T* Across Multiple Floral Morphs

The expression levels of *KFB^T* in various morphs of *P. vulgaris* flowers were quantified via RNA-Seq analysis. The floral morphs were pin, thrum, a short homostyle mutant and two lines of long homostyle mutants (Li *et al.*, 2016; Crosby, 1940; Crosby, 1949). The *S* locus is absent from the pin genome but this sample was included as a negative control to prove that expression does not occur. The flowers present different morphologies and this experiment was used to potentially connect these morphologies with *KFB^T* expression and provide early indications of its function in heteromorphy.

All five *Primula* lines were from the lab glasshouse population (Chapter 2.1). Both long homostyle lines were derived from UK populations naturally occurring in the wild, one from Somerset (Crosby, 1940; Li *et al.*, 2016) and one from the Chiltern Hills (Crosby, 1949; Li *et al.*, 2016). Mature flower buds were sampled prior to opening at a length of 15-20 mm by Dr Jinhong Li. Four biological replicates were used for each morph. RNA was extracted by Dr Li and Illumina sequenced as paired-ends by the Platforms & Pipelines team at the Earlham Institute, Norwich.

3.7.1 Mapping RNA-Seq Reads to the *P. vulgaris* Genome

Before expression levels of *KFB^T* in the sample transcriptomes could be quantified, the raw sequencing reads had to be mapped to the *P. vulgaris* genome. The long homostyle assembly by Dr Jonathan Cocker was used as the template because it is the highest quality *P. vulgaris* genome available (Li *et al.*, 2016). Contigs smaller than 200 bp had previously been filtered from the genome, which was queried against a library of known contaminants to remove such sequences and improve its quality (Li *et al.*, 2016). The twelve homostyle RNA-Seq samples (containing four biological repeats for each of the two long and one short homostyle mutant lines) were individually sequenced on two lanes, meaning there was a technical repeat for each. The four biological repeats for each pin and thrum sample were sequenced at an earlier date on just one lane (Li *et al.*, 2016).

Version 2.1.0 of the 'Hierarchical Indexing for Spliced Alignment of Transcripts' (HISAT; Kim *et al.*, 2015) package was used for alignment of the RNA-Seq reads to the *P. vulgaris* genome. Using a splice-aware tool was essential due to the alignment of RNA to a DNA template, therefore it was imperative that the algorithm accommodated gaps caused by the removal of introns. The `hisat2-build` command was used to make an index of the

reference genome, as required by the HISAT tool (Kim *et al.*, 2015). The `hisat2` command was subsequently used with each set of paired read files for each sample to begin alignment of the RNA-Seq reads to the reference genome. The input files were in FASTQ format so the `-q` argument was used with this command. The `-x` argument was used because paired reads were being provided as separate files and the `--dta` argument was used, as stipulated by the StringTie package (Pertea *et al.*, 2016) which was used for downstream analysis.

3.7.2 Quantification of *KFB^T* in Multiple Morphs of *P. vulgaris* Flowers

To quantify expression levels of *KFB^T* in the various morphs that underwent RNA-Seq, the number of transcripts that mapped to this gene were enumerated. The output from the HISAT tool consisted of a Sequence Alignment Map (SAM) for each sample. The `samtools depth` command was used with the `-aa` argument on these files to display the coverage for every base in the genome; including those to which zero transcripts mapped.

The *P. vulgaris* reference genome (Li *et al.*, 2016) was queried with the *KFB^T* coding region to locate the exact scaffold and position at which the gene is located. The mean coverage was then calculated across these co-ordinates for each depth file, to quantify the average expression levels of *KFB^T* in each sample. The number of raw reads was deduced from the RNA-Seq files so that the calculated expression levels could be converted into figures of 'transcripts per million' (TPM), which is the standard way to present normalised RNA-Seq data. The mean TPM values and standard errors were calculated across the four biological repeats for each floral morph and presented as a bar chart (Figure 3.7.2).

The analysis was repeated in three different ways to check consistency across the results. Average TPM values from the StringTie (Pertea *et al.*, 2016) output across the *KFB^T* region were calculated and presented on a bar chart. A second check used the average read counts from StringTie instead. The final check involved using version 0.44 of the Kallisto package (pachterlab.github.io/kallisto/) to align and quantify the RNA-Seq reads against the *KFB^T* gene (not the entire genome). Expression level patterns were consistent across all four methods, which confirmed authenticity of the results.

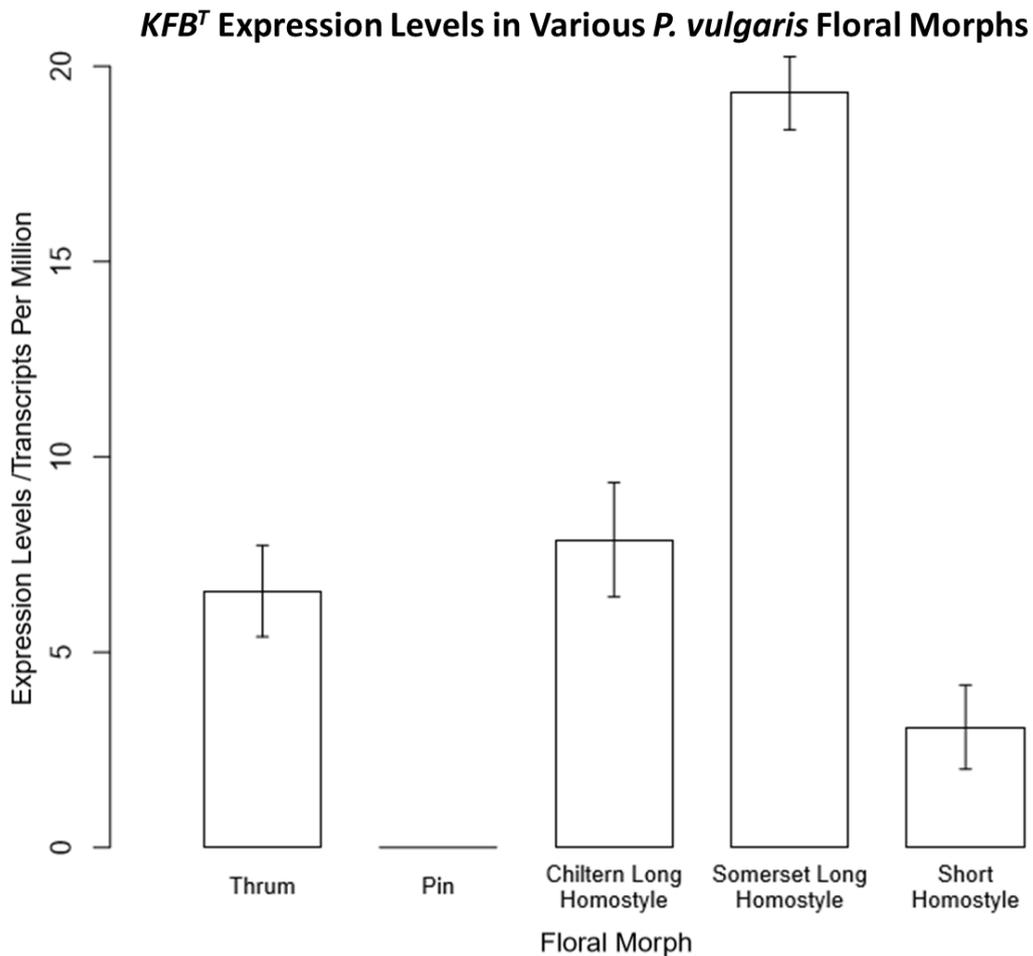


Figure 3.7.2: Quantification of *KFB^T* expression across thrum, pin, short homostyle, Chiltern long homostyle and Somerset long homostyle flowers via RNA-Seq. Error bars represent standard error.

The Somerset long homostyle morphs presented a mean 2.94-fold increase in *KFB^T* transcription when compared to thrum, however the Chiltern long homostyle and thrum samples displayed similar levels of *KFB^T* expression (Figure 3.7.2). There are no known phenotypic differences between long homostyles of the Chiltern and Somerset populations. Short homostyle flowers exhibited a mean 2.13-fold decrease in *KFB^T* expression levels when compared to thrum (Figure 3.7.2). The short homostyle phenotype is linked to a mutation in *GLO^T* that causes a reduction in anther height (Li *et al.*, 2016). The pin flower presents no *KFB^T* transcripts because this gene is part of the *S* locus, which is absent from the pin genome (Li *et al.*, 2016).

3.8 Differential Expression of *P. vulgaris* Genes In Multiple Floral Morphs

RNA-Seq data had been used to specifically quantify *KFB^T* expression in flowers of thrum, pin, short homostyle, Chiltern long homostyle and Somerset long homostyle (Chapter 3.7). However, the dataset also contained transcriptome-wide information. The long homostyle phenotype is caused by a mutation in *CYP^T* (Huu *et al.*, 2016; Li *et al.*, 2016) and the short homostyle phenotype is linked to a *GLO^T* mutation; both of which lose self-incompatibility (Li *et al.*, 2016). The morphology of pin flowers is caused by absence of the entire *S* locus (Li *et al.*, 2016).

The phenotypes presented by these flowers are expected to be a result of differential gene expression. It was therefore hypothesised that elucidating differentially expressed genes between these RNA-seq samples could identify candidates involved in the networks that ultimately regulate these various physiologies. This information would offer insight regarding the end-to-end pathway that underpins both floral heteromorphy and self-incompatibility in *Primula*. It was also important to establish this dataset in advance of a planned yeast two-hybrid experiment (Chapter 6) to query for potential differential expression of genes encoding any identified *KFB^T* partner proteins, which could offer significant support for functional models of *KFB^T*.

StringTie (Pertea *et al.*, 2016) was used in Chapter 3.7.2 to map RNA-Seq reads from each sample to the *P. vulgaris* genome. These output files also contained normalised expression levels for each predicted transcript, in the form of Transcripts Per Million (TPM) values. A shell script for the Linux Command Line was written (Appendix A7) that utilised the General Feature Format file from the *P. vulgaris* genome annotation with the StringTie output files for each sample to calculate the average expression levels of every gene.

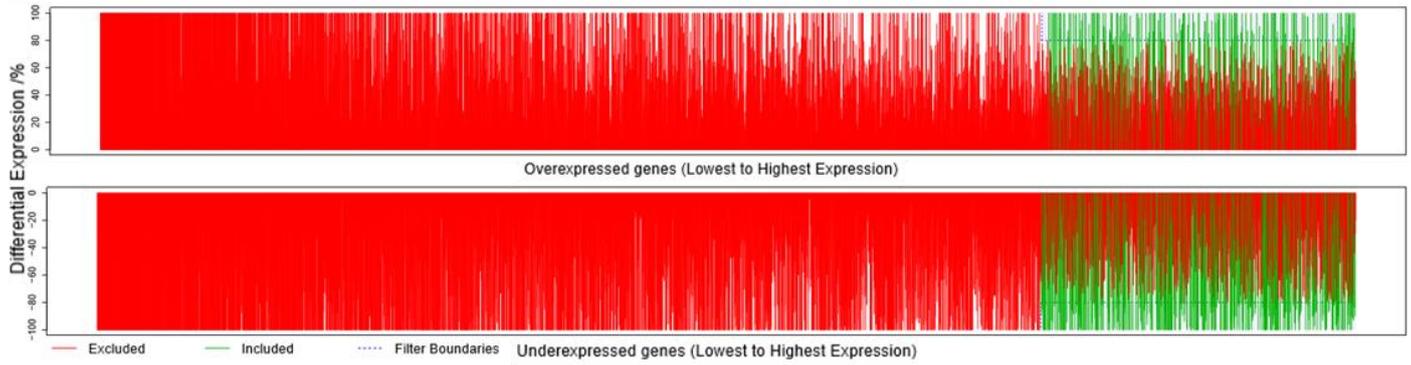
With reference to wildtype thrum, overexpressed genes in each sample were separated from under-expressed genes. Genes were first sorted in order of their expression levels. The difference between thrum TPM and corresponding sample TPM value was calculated for each gene. These differential expression figures were either converted to a percentage of the sample TPM for overexpressed genes or a percentage of the thrum TPM for underexpressed genes (i.e. the larger of the two figures). Values for under-expressed genes were multiplied by -1.

Differential expression percentages were presented on bar charts for each sample (Figure 3.8-1). As genes were first ordered by their raw TPM values, bars further to the right were

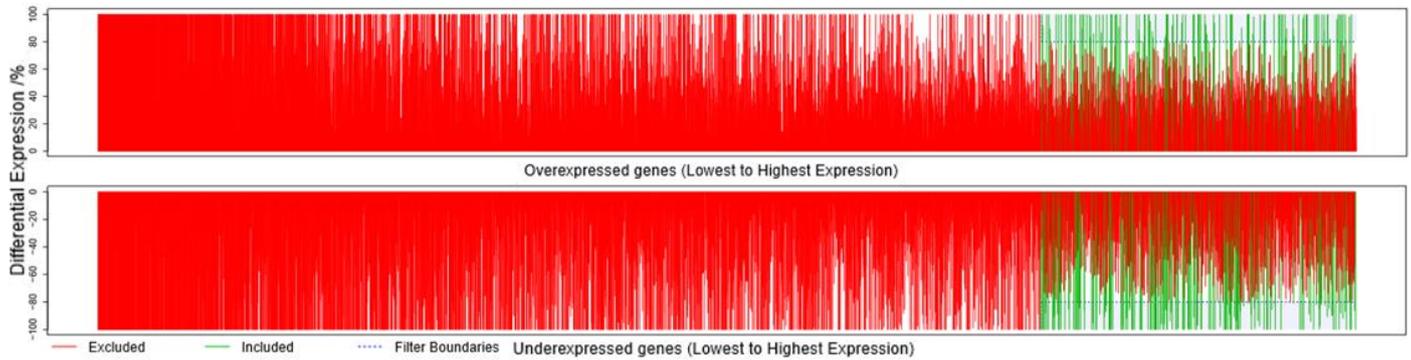
deemed to exhibit more significant differential expression. For example, a 100 % bar on the right of the overexpression graph could mean 6000 transcripts in the sample flower versus 0 transcripts in the thrum, whereas the same bar on the left of the plot could represent just 1 transcript in the sample against 0 in thrum. The bars were filtered to only include those on the righthand quarter of the graph with differential expression figures that exceeded 80 %. These thresholds were arbitrary but presented a high-priority list of candidates for further investigation.

Differential Gene Expression Between *P. vulgaris* Thrum and Various Floral Morphs

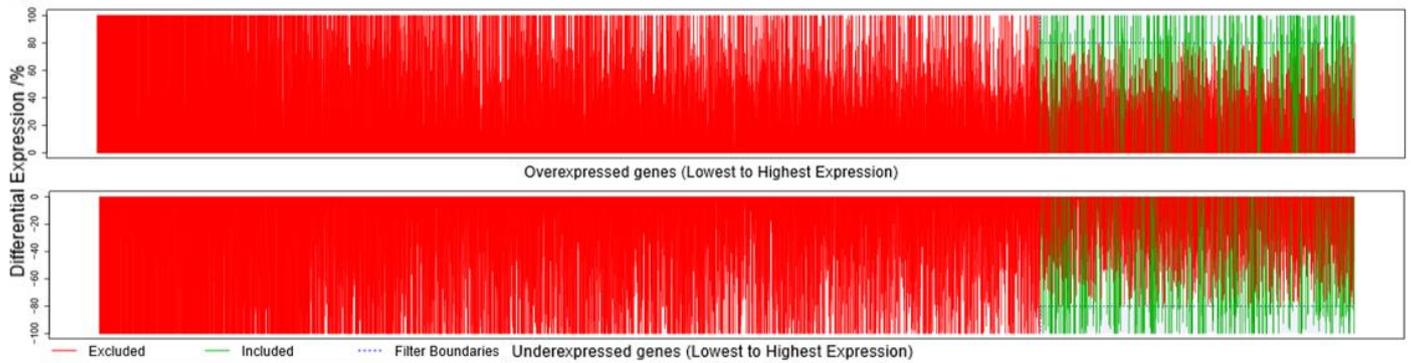
Short Homostyle



Chiltern Long Homostyle



Somerset Long Homostyle



Pin

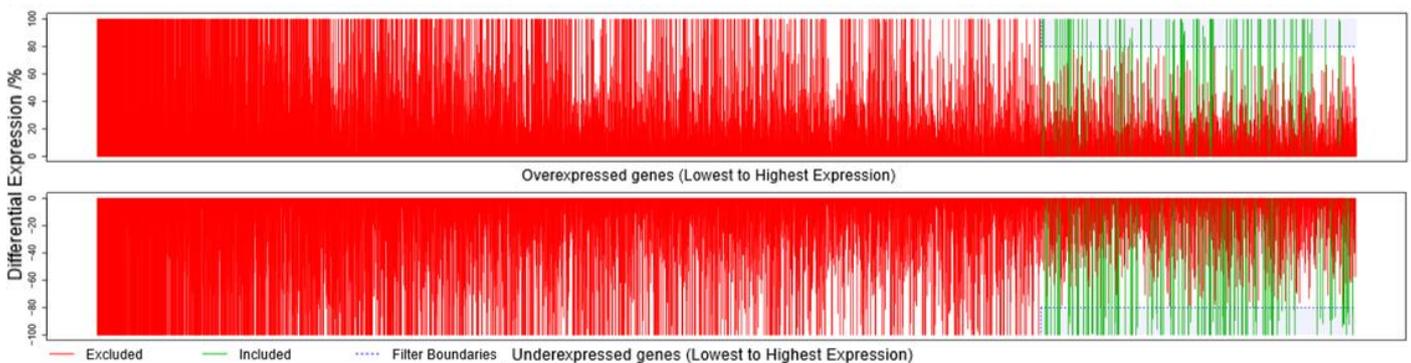


Figure 3.8-1: RNA-Seq data was used to estimate average expression levels for every gene from five *P. vulgaris* floral morphs. After ordering genes by normalised transcript levels, differential expression was calculated between thrum and the other morphs before filtering. The most significant 25 % of genes with >80 % differential expression were extracted for further analysis (green). The others were discarded (red).

Resultant gene lists from each sample were compared to identify common entries (Figure 3.8-2). For example, genes common to the Chiltern long homostyle, Somerset long homostyle and pin sample lists could be indicative of differential expression that occurs when *CYP^T* is disrupted by mutation (Huu *et al.*, 2016; Li *et al.*, 2016). Likewise, genes common to the short homostyle and pin lists may be affected by *GLO^T* absence (Li *et al.*, 2016). Genes in only the long or short homostyle lists could be candidates for involvement in the self-incompatibility pathway, which is broken down in these mutants.

The Somerset and Chiltern long homostyle plants each respectively presented 271 and 139 overexpressed genes for further investigation, of which 65 were common to both lists (Figure 3.8-2). These were carried forward and 19 were found to be in common with the overexpressed genes identified in pin flowers (Figure 3.8-2). Both long homostyle and pin plants have non-functional or absent *CYP^T* genes that fail to suppress anther height (Huu *et al.*, 2016).

The *GLO^T* gene is dysfunctional in short homostyles and entirely absent from the pin genome (Li *et al.*, 2016). Short homostyle and pin samples were found to share 33 overexpressed genes (Figure 3.8-2). Of the 65 overexpressed long homostyle genes, 40 were found to be in common with those from short homostyle flowers (Figure 3.8-2). Mutations present in these homostyle genomes leads to a breakdown in the self-incompatibility system.

There were more under-expressed than overexpressed genes identified (Figure 3.8-2). This is true for every sample except the Somerset long homostyle, though there was still an increased number of under-expressed genes in common with the Chiltern long homostyle (Figure 3.8-2). Of these 116 long homostyle genes, 32 were shared with under-expressed genes identified in pin and 90 were shared with those from short homostyle flowers. A total of 73 under-expressed genes were found in common between the short homostyle and pin morphs. There were 14 overexpressed and 25 under-expressed genes common to all three tested morphs (Figure 3.8-2).

Number of Differentially Expressed Genes in Various *P. vulgaris* Flowers Compared to Thrum

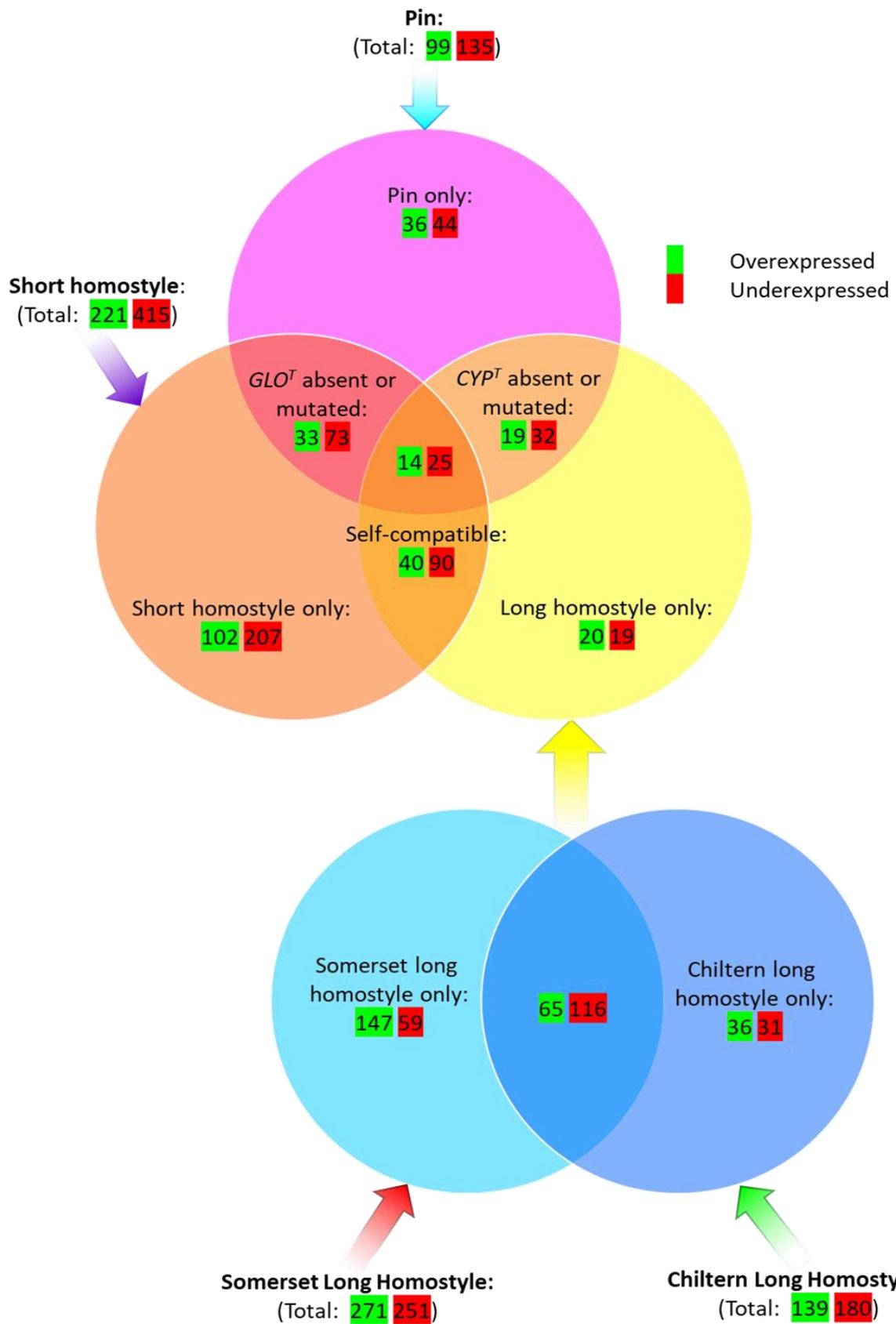


Figure 3.8-2: Number of most significantly upregulated and downregulated genes in RNA-Seq datasets from flowers of thrum, pin, short homostyle and two long homostyle populations.

Transcript levels of nineteen genes that were deemed to exhibit the most significant differential expression in at least two morphs are shown in Figure 3.8-3. Their corresponding descriptions were obtained via BLASTx and presented in Table 3.8. This identified two genes encoding lipid transfer proteins as well as two genes involved in the transport of UDP. The plant self-incompatibility S1 gene from *Corchorus capsularis* should also be a top priority for further investigation. Transcription of this gene appears to be significantly upregulated in pin and long homostyle plants, which both lack a functional *CYP7* gene (Huu *et al.*, 2016; Li *et al.*, 2016).

Differentially Expressed Gene Candidates Between *P. vulgaris* Floral Morphs

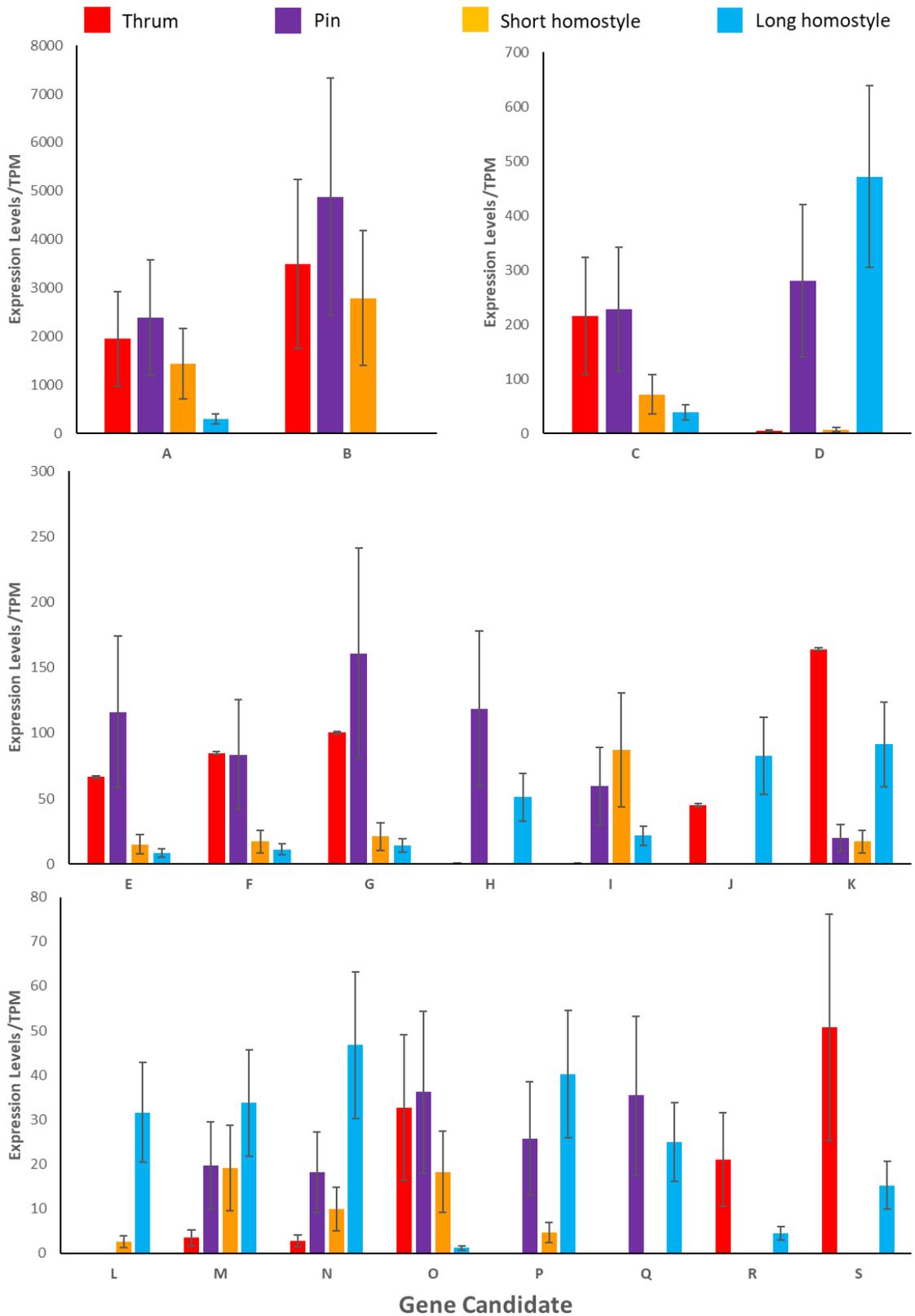


Figure 3.8-3: Transcript levels of genes predicted to exhibit the most significant differential expression between *P. vulgaris* flower morphs of thrum, pin, short homostyle and long homostyle. Only those that demonstrated differential expression across at least two groups were included.

Table 3.8: BLASTx results to identify candidate genes that exhibit greatest differential expression between thrum, pin, short homostyle and long homostyle *P. vulgaris* flowers.

	Description	Species	E-value	Accession
A	Non-specific lipid-transfer protein	<i>Prunus yedoensis</i>	5e-26	PQP98495.1
B	Non-specific lipid-transfer protein 1-like	<i>Quercus suber</i>	3e-33	XP_023900811.1
C	Thiamine thiazole synthase	<i>Coffea arabica</i>	0	XP_027064965.1
D	Plant self-incompatibility S1	<i>Corchorus capsularis</i>	2e-04	OMO53010.1
E	UDP-N-acetylglucosamine transferase	<i>Camellia sinensis</i>	2e-91	XP_028100452.1
F	Histone H2A.1-like	<i>Lactuca sativa</i>	2e-76	XP_023755243.1
G	Calvin cycle protein like	<i>Actinidia chinensis</i>	3e-40	PSS14728.1
H	Hypersensitive-induced response protein	<i>Juglans regia</i>	0	XP_018831584.1
I	Transmembrane EMP24 domain-containing protein	<i>Coffea arabica</i>	7e-95	XP_027098355.1
J	DnaJ protein homolog	<i>Ananas comosus</i>	0	XP_020092053.1
K	Cation/H(+) antiporter 19	<i>Eucalyptus grandis</i>	0	XP_010043343.1
L	UDP-galactose/UDP-glucose transporter	<i>Actinidia chinensis</i>	0	PSS13978.1
M	Integrin-linked protein kinase 1-like	<i>Camellia sinensis</i>	0	XP_028062147.1
N	F-box protein	<i>Actinidia chinensis</i>	0	PSS29169.1
O	Probable carotenoid cleavage dioxygenase 4 chloroplastic	<i>Vitis vinifera</i>	3e-54	XP_002269538.2
P	Subtilisin-like protease	<i>Actinidia chinensis</i>	0	PSS216061.1
Q	Vacuolar protein sorting-associated 8 homolog	<i>Camellia sinensis</i>	2e-34	XP_028073451.1
R	Probable L-ascorbate peroxidase 6	<i>Glycine max</i>	0	XP_003523435.2
S	Oxygen-evolving enhancer protein	<i>Parasponia andersonii</i>	7e-103	PON31702.1

3.9 Discussion

3.9.1 Comparison of KFB^T Against *Arabidopsis* KMD Homologues

Comparisons were made between KFB^T and homologous proteins with known function to gain early insight into potential roles within *P. vulgaris*. The KFB^T amino acid sequence was aligned against the Kiss me Deadly (KMD) proteins from *Arabidopsis*; its four closest homologues (Figure 3.5 upper). It was found to be more similar to both KMD1 and KMD2 than KMD3 or KMD4 (Table 3.3.5). Although longer sequences would be expected to have an increased chance of containing matches during alignment, KMD1 and KMD2 were the shorter of the two pairs by approximately 15 %. The *KMD1* and *KMD2* genes are on the same *Arabidopsis* chromosome – on opposite sides of the centromere – and *KMD3* and *KMD4* are near the telomeric side of two different chromosome arms (Sun *et al.*, 2007).

Of the four KMD proteins, KMD1 and KMD2 have been shown to interact with the *Arabidopsis* SKP1 (ASK1) subunit of the SCF complex responsible for protein degradation but KMD4 did not (Zhang *et al.*, 2013). Conversely, binding between all four KMD proteins and ASK1 has elsewhere been confirmed (Kim *et al.*, 2013a). While these studies present conflicted findings, they both agree on the affinity between ASK1 with KMD1 and KMD2, which together displayed most similarity to KFB^T. Expression of *KMD1* and *KMD2* was most abundant at the shoot meristem in *Arabidopsis*, with *KMD2* and *KMD3* accumulation in the root meristem too (Kim *et al.*, 2013a). Root expression of *KMD2* was the highest of all 97 Kelch F-box genes in *Arabidopsis*, demonstrating an approximate 3-fold increase over all but one of them (Sun *et al.*, 2007). Both areas rich in *KMD2* transcripts are active regions of cell division.

The KFB^T amino acid sequence shared greatest similarity with that of KMD2, which was the *Arabidopsis* KMD found to exhibit greater expression in flowers and buds than other tissues (Zhang *et al.*, 2013). This study did not examine *KMD3* but *KMD2* displayed the lowest average expression compared to the other two. Conversely, Sun (*et al.*, 2007) found *KMD3* to have the highest expression of all 97 *Arabidopsis* Kelch F-box genes in young flower tissue by approximately double in comparison to those outside of the *KMD* family. Moreover, *KMD4* exhibited second highest floral expression and approximately 3-fold greater anther expression than all other *Arabidopsis* Kelch F-box genes. The former experiment utilised quantitative PCR and the latter employed a microarray method, which possibly explains disparity between their findings.

Further differential expression analyses uncovered reduced *KMD2* transcription in sunflower embryos treated with methylviologen, which inferred a role for this gene in abscisic acid signalling (El-Maarouf-Bouteau *et al.*, 2014). In *Arabidopsis* grown in reduced gravity on the International Space Station, a 1.5-fold increase in *KMD2* expression was observed alongside 1.28-fold and 1.23-fold increases of *KMD3* and *KMD1*, respectively (Weitzel *et al.*, 2016). The overexpression of *KMD2* was linked to lightening of the UV-protectant seed coat in *Arabidopsis* and the KMD proteins were subsequently shown to regulate phenylpropanoid biosynthesis by degradation of four phenylalanine ammonia-lyase enzymes (Zhang *et al.*, 2013).

In a separate experiment, *KMD2* was the only one of the four *Arabidopsis* homologues not to undergo transcriptional reduction when plants were exposed to cytokinin for 1 hr (Kim *et al.*, 2013a), although in the same year the authors found *KMD2* to downregulate cytokinin response in rice (Kim *et al.*, 2013c). Unlike the other three *KMD* genes, *KMD2* was not found to demonstrate circadian dependence in response to environmental factors (Kim *et al.*, 2013a).

Functional redundancy amongst the KMD family further impedes speculation regarding the role of KFB^T. For example, the lignin content of *Arabidopsis* plants has been shown to reduce when either *KMD1*, *KMD2* or *KMD4* are overexpressed (Zhang, 2013). The knockdown of *KMD3* transcripts has also been accomplished by using a *KMD4* antisense construct in *Arabidopsis* (Kim *et al.*, 2013a).

Although KFB^T has displayed greatest similarity to what appears to be the more flower-specific of the KMD homologues (Zhang *et al.*, 2013), the vast number of diverse roles for these proteins has made it difficult to establish refined hypotheses regarding KFB^T function. However, it is reasonable to expect that KFB^T may contain an F-box domain for interaction with SKP1 of the ubiquitinase pathway in *Primula*. As roles in both cytokinin response and the abscisic acid pathway have been discussed in literature, KFB^T could likely have a role in modulating sensitivity to plant hormones. Its method of action may also be indirect, as seen in the targeting of the phenylalanine ammonia-lyase enzyme to limit phenylpropanoid biosynthesis (Zhang *et al.*, 2013) and targeting of transcription factors to modulate cytokinin response in *Arabidopsis* (Kim *et al.*, 2013b).

If KFB^T can recruit numerous proteins for degradation – like *KMD2* – then target specificity must be conveyed by tight regulation of its spatial and temporal expression instead. It

follows that seemingly functionally redundant Kelch F-box genes have demonstrated differential expression, which so far remains unexplained (Schumann *et al.*, 2011). Increased impetus was therefore placed on analysis of regulatory promoter elements (Chapter 3.6) and qPCR experiments (Chapter 4).

3.9.2 Characterisation of Kelch Repeat and F-Box Domains in *KFB^T*

Kelch proteins can be phylogenetically grouped based on their number of Kelch repeats (Schumann *et al.*, 2011). Kelch proteins without F-box domains usually contain five to seven repeats (Li *et al.*, 2004) but Kelch F-box proteins in plants contain between one and five (Schumann *et al.*, 2011). Of the 263 Kelch proteins tested in *Arabidopsis*, 14 % contained a single Kelch repeat, 59 % had two, 19 % possessed three, 4 % had four and the final 4 % contained five (Sun *et al.*, 2007). Analysis of the entire Pfam database found 65 % of Kelch proteins to only contain one repeat (Schumann *et al.*, 2011). The author speculated how the propeller structure could be completed with just one Kelch motif but concluded that these proteins were likely to contain further repeats that simply went undetected by the bioinformatical tools due to degradation of the Kelch consensus toward the C-terminus.

Five Kelch repeats were identified in *KFB^T*. Unlike the gradual C-terminal degradation of Kelch motifs observed by Schumann (*et al.*, 2011), the fifth *KFB^T* Kelch repeat was a closer match than the first (Figure 3.1). These motifs did not exhibit less coherence with the consensus sequence toward the C-terminus of the protein. The fourth and fifth repeats were separated by a 29 amino acid spacer that appeared to contain half a Kelch motif. Fractional Kelch repeats have been identified in other proteins (Kutuzov *et al.*, 2002). Only two or three Kelch repeats have been identified in the AtKMD proteins (Zhang *et al.*, 2013). Alignments of the *KFB^T* amino acid sequence to four *Arabidopsis* KMD proteins, 154 *P. vulgaris* Kelch proteins and *KFB^T* sequences from sixteen *Primulaceae* family members (Figure 3.5) all presented a central region of high similarity that directly corresponded to the middle three Kelch repeats in *KFB^T* (Figure 3.1). The region between the fourth and fifth Kelch motifs was highly variable and similarity also reduced over the fifth Kelch repeat but significantly increased toward their C-termini. While this region has not yet been associated with a functional domain, it has been proposed that the C-terminus of proteins containing Kelch domains may fold to play important roles by capping the propeller and protecting the hydrophobic core (Li *et al.*, 2004). A truncated Kelch consensus has even

been found to commence prior to a stop codon at the 3'-end of the human *Keap1* gene yet the remainder of the blade resumed after the start codon (Li *et al.*, 2004) – though this did not encode an F-box protein.

A string of 45 amino acids precedes the Kelch domain in KFB^T . The first 48 amino acids of the *Arabidopsis* KMD proteins were found to contain the F-box domain (Kim *et al.*, 2013a). The F-box consensus sequence used by the authors found six matches, nine highly similar hits and seventeen weak matches to the KMD proteins. This same consensus sequence identified ten matches, eight highly similar hits and six weak matches to the N-terminal region of KFB^T (not shown). Findings by Kim (*et al.*, 2013A) thus support these position, length and consensus sequence analyses to confirm presence and location of the F-box domain in KFB^T . This explains the highly similar N-termini consistently observed across alignments in which no Kelch repeats were located (Figures 3.5 & 3.1).

While attempts have been made to elucidate an F-box consensus sequence, these domains are notoriously difficult to identify due to the low number of consistent positions (Kipreos & Pagano, 2000). Similarity between the F-box regions of KFB^T and other proteins was perceived to decline toward the C-terminal Kelch domain (Figure 3.5). Divergence at the C-terminus of F-box sequences has elsewhere been observed (Mercer *et al.*, 2005).

Although Kelch domains on different *P. vulgaris* proteins have different target substrates, the F-box would be expected to bind with the same partner protein: an SKP1-like subunit of an SCF complex, of which there are approximately nineteen variants in *Arabidopsis* (Takahashi *et al.*, 2004). However, this N-terminal region of similarity was not as wide between the 155 *P. vulgaris* Kelch proteins than in the other two alignments. This is because not every member of the Kelch protein family has an N-terminal F-box domain. There were also far more samples compared in this alignment, which inevitably introduces much greater sequence diversity.

The number of synonymous to non-synonymous mutations across the KFB^T sequences from sixteen *Primulaceae* species were tallied (Figures 3.3.5-1 & 3.3.5-3). This is an indicator of the direction of evolutionary selection pressure toward either divergence or purification (Zhang *et al.*, 2006). The average number of non-synonymous mutations increased toward the 5'-end of the F-box consensus sequence identified in KFB^T , reinforcing the earlier point of 5'-end variance in F-box sequences. However, a clear crossover was observed between the two mutation types and a raised ratio of synonymous to non-synonymous substitutions

was contained to the central three Kelch repeats, suggesting their evolutionary resistance to change. The most conserved region was across the second Kelch motif and crystal structure analysis of the human Keap1 protein showed that extra residues at this second propeller blade form protrusions that directly interact with the recruited target protein (Li *et al.*, 2004). Preservation of this positively charged structure was so imperative that mutations here resulted in total loss of binding affinity to the substrate (Li *et al.*, 2004). The mutation comparisons (Figure 3.3.5-3) show a strongly diversifying region across the fifth KFB^T Kelch repeat and the preceding space to the fourth motif, in which a fractional Kelch repeat was located. The graph returns to purifying selection at the C-terminus. This latter region of conservation reinforces an earlier proposition that the C-terminus may present a structure important in closing the β -barrel propeller formed by the Kelch domain.

3.9.3 Identification of *P. vulgaris* Kelch Protein Family

Amino acid sequences from the *P. vulgaris* genome were queried for other proteins potentially containing Kelch domains. Kelch motifs are usually 44 to 56 amino acids long and share as little as 11 % similarity between different proteins (Bork & Doolittle, 1994), which makes them difficult to identify via consensus sequence. Kelch motifs at the same position in different proteins have been found to share more similarity than different Kelch motifs within the same protein and this suggests each blade may have a particular role in the final folded propeller (Schumann *et al.*, 2011).

The motif search parameters therefore had to be designed around common landmarks of the Kelch consensus sequence (Figure 3.2-1). A diglycine pair, leucine and tryptophan are conserved in over 90 % of Kelch sequences, with the leucine and tryptophan residing a distance of six amino acids apart in approximately 70 % of motifs (Adams *et al.*, 2000). The *P. vulgaris* search parameters permitted this leucine to be replaced by either phenylalanine or tyrosine instead. The double glycine pair is important for the sharp turn of the protein propeller structure and is separated from the downstream leucine by a linker of varying length (Li *et al.*, 2004).

Common linkers of 11 to 19 (Adams *et al.*, 2000) and 12 (Andrade *et al.*, 2001) were noted in the literature. A gap of 9 to 25 amino acids was implemented in the *P. vulgaris* screen. Smaller gaps do exist but, although reducing this lower parameter boundary would have included more Kelch proteins (Figure 3.2-2 upper), the primary aim of this investigation was to identify KFB^T-like proteins and not to exhaustively define the entire Kelch family.

Another 32 Kelch candidates would have been obtained if this lower gap boundary parameter had been reduced to nine (not shown). Moreover, candidates were filtered out that did not contain at least two Kelch repeats, though proteins containing a single motif have been identified in plants (Schumann *et al.*, 2011). The number of identified genes would have risen to 1320 (not shown) if parameters had been adjusted to include single motif Kelch candidates.

Claims of 96, 97 and 103 Kelch F-box proteins in *Arabidopsis* have been made (Zu *et al.*, 2009; Sun *et al.*, 2007; Schumann *et al.*, 2011). A further 36 were identified in *Vitis vinifera*, 68 in *Populus trichocarpa*, 39 in *Oryza sativa*, 44 in *Sorghum bicolor*, 46 in *Selaginella moellendorffii* and 71 in *Physcomitrella patens* (Schuman *et al.*, 2011). On the contrary, Zu (*et al.*, 2009) identified only 35 in poplar and 27 in rice. This highlights the difficulty in defining such variable proteins. Just three Kelch F-box proteins were identified in walnut and they demonstrated greater expression in flowers than nutritive tissues (Yan *et al.*, 2019).

A total of 155 Kelch proteins were identified in *P. vulgaris*, including KFB^T itself (Figure 3.2-2 upper). Two of these demonstrated markedly increased similarity to KFB^T (Figure 3.2-3), one of which had an amino acid sequence of exactly the same length and its most significant BLAST hit was to KMD3 (Table 2.11.1). The location of this candidate in the genome was checked to see whether it was near any other genes similar to those in the *S* locus to perhaps begin understanding how this tightly linked supergene was derived, but no significant findings were made.

Of the top fifteen most similar *P. vulgaris* Kelch candidates to KFB^T, ten of them were confirmed as F-box/Kelch repeat proteins by BLAST and another was clearly linked to the ubiquitinase pathway in which KFB^T has expected involvement (Table 2.11.1). A phylogenetic tree from Clustal alignments was also analysed (Figure 3.2-3) but the irrelevance of its closest hits to KFB^T suggested superiority of the former analytical method.

The most commonly occurring amino acids across alignments of these *P. vulgaris* candidates were landmarks of the Kelch consensus (not shown) but little significance should be appointed to this as the dataset was biased toward these landmarks; sequences without them would have been filtered out by the search parameter criteria. The F-box binds with the SCF complex and the Kelch domain confers specificity to the substrate (Gupta & Beggs, 2014). As members of the Kelch family would be expected to recruit

different target proteins for degradation, information about the KFB^T target could have been uncovered by alignments with the other Kelch candidates. However it appears that prevailing similarity across the Kelch domain masks any subtler differences that would be indicative of substrate specificity (Figure 3.5 centre). This is likely because most of the domain is devoted to preserving the important architecture of the complex propeller structure and only a small region of it is responsible for target affinity.

It was noted that six pairs of the 155 *P. vulgaris* Kelch proteins were adjacent in the genome. The tendency of Kelch genes to gather in adjacent clusters has been noted elsewhere (Andrade *et al.*, 2001). Plant Kelch F-box proteins have been organised into eighteen subfamilies and those in group 5 are commonly found as tandem repeats (Sun *et al.*, 2007). The *KMD1* gene is in this family. It is believed that 38 of the 65 *Arabidopsis* Kelch F-box proteins resulted from tandem duplication (Sun *et al.*, 2007). Schumann (*et al.*, 2011) found a link between gene stability and spatial position on the genome, noting that superstable genes were evenly dispersed across chromosomes while unstable ones were strongly clustered. The similarity and intervening distances of the six adjacent *P. vulgaris* Kelch pairs were compared against those of other genes (Figures 3.2-4 & 3.2-5). It is believed that the varying similarities between these pairs correlates to the time since their gene duplication events.

3.9.4 Conservation of KFB^T Across the *Primulaceae* Family

The presence of KFB^T in genomes of all sixteen tested *Primulaceae* species was confirmed (Figure 3.3.4). This is not only suggestive of the importance of KFB^T in floral heteromorphy but also supports applicability of the current *S* locus model across the *Primulaceae* family. Phylogeny from non-coding chromosomal DNA sequences placed a clade containing *P. incana* and *P. laurentiana* alongside a clade containing *P. halleri* and *P. scotica* with *P. yuparensis* positioned independently (Mast *et al.*, 2001). Alternatively, the Clustal alignment in Figure 3.3.5-2 placed *P. halleri* and *P. incana* together in a clade alongside a separate clade containing *P. laurentiana* and *P. yuparensis* with *P. scotica* positioned independently instead. Both the former and latter phylogenies are reflected in Figure 3.3.4 and appear to be distinguished by whether nucleotide (upper) or amino acid sequences (lower) are aligned.

Although *P. vulgaris* and *P. veris* both have eleven chromosomes, they appear to be more dissimilar from species with the same number of chromosomes (*P. prolifera*, *P. prenantha*,

P. chungensis, *P. cuneifolia* and *P. cockburniana*) than those with nine (*P. yuparensis*, *P. laurentiana*, *P. incana*, *P. helleri* and *P. scotica*) according to phylogeny proposed by Mast (*et al.*, 2001). The *P. kewensis* sample is also in this group and resulted from hybridisation of *P. floribunda* and *P. verticillata*, which both have nine chromosomes. The most dissimilar species was *Hottonia palustris*. This has a chromosome number of ten and was also the only tested species in which the *KFB^T* start codon was mutated. It is unclear whether this gene is thereby disabled or if the consecutive start codon 66 bp upstream is active. This start codon is in the correct frame but the extra 22 amino acids would significantly modify the F-box domain believed to exist at the N-terminus and may render the protein dysfunctional. No differing phenotypic patterns could be found between these species that would infer a role for *KFB^T*.

The aforementioned group of *Primula* species with nine chromosomes all presented the lowest number of mutations against a consensus sequence made from the sixteen *Primulaceae KFB^T* homologues. However, this could simply mean they are very closely related and therefore skewed the consensus sequence in their own favour. The synonymous to non-synonymous substitution ratios of *P. chungensis*, *P. cockburniana* and *P. prenantha* indicates these species are the most likely to be diverging, which suggests they contain advantageous mutations undergoing positive selection. As non-synonymous mutations occur less frequently than synonymous substitutions (Nekrutenko *et al.*, 2002), ratios greater than 1 for these species are probably underrepresented and should perhaps be emphasised.

This analysis was inspired by K_a/K_s calculations that represent the direction of evolutionary selection pressure to favour or resist change. It would be desirable to redo this analysis across the 155 Kelch proteins identified in *P. vulgaris* but their sequences are of such differing lengths and dissimilar content that obtaining a fair consensus across them all would be impossible.

The method used here compared each codon from a sample to the corresponding codon in a consensus sequence derived from all sixteen *Primulaceae KFB^T* nucleotide sequences. If a codon differed, each base was tested to see whether it would modify the amino acid and, in this way, the synonymous and non-synonymous substitutions were recorded. These figures should not directly be compared against K_a/K_s values from other publications because those are commonly the product of more complex equations in which synonymous or non-synonymous substitutions are enumerated per synonymous or non-synonymous

site. These raw values subsequently undergo statistics to weight transversions against transitions and account for multiple substitutions.

The novel script written for this *Primulaceae* analysis was sufficiently intricate to account for all information without loss through normalisation of the figures. One potential issue for future use would be oversensitivity of ratios in sequences possessing a low amount of mutations. For example, a sample could present a high non-synonymous to synonymous substitution ratio of 2 (suggestive of evolutionary divergence) even if it had only acquired three mutations, whereas a far more heavily mutated sample could present a lower ratio due to a more balanced mixture between the two substitution types. Such problems were here circumnavigated by presenting the ratios in context by using stacked bar plots (Figure 3.3.5-1). These non-normalised raw figures were also useful for the line chart in Figure 3.3.5-3, which proved to be very informative regarding preservation and divergence across domains of the *KFB^T* nucleotide sequence.

The mechanism by which *S* locus integrity is maintained remains unknown. The hemizygous nature of this gene cluster means DNA repair mechanisms that depend on the homologous chromosome for use as template (such as one model of double stranded break repair) must be unable to function. Despite this, accumulation of mutations in *KFB^T* appears to be low across the *Primulaceae* family (Figure 3.3.5-1) and these genes demonstrated an average reduction in similarity of just 4.53 % when nucleotides were aligned instead of amino acid sequences (Figure 3.3.4). The clear purifying selection across the Kelch region (Figure 3.3.5-3) proves that reparation mechanisms must somehow take place. It would therefore be of great interest to use sequence capture and repeat the synonymous substitution analyses on a gene from outside the *S* locus to compare their mutation rates and selection pressures. This data is not available at present.

3.9.5 Transcriptional Regulation of *KFB^T*

The upstream region of the *KFB^T* nucleotide sequence was screened for common promoter elements to try and identify potential binding sites for transcription factors and other regulatory features. Firm conclusions were hampered by variation in the consensus sequences of these elements.

There is a 1 in 512 chance of finding a 5 bp fragment coherent with the Downstream Promoter Element consensus sequence ([A/G]G[A/T][C/T][G/A/C]) but 35 potential matches

were found in the ~3 kb upstream region of *KFB^T*. A TATA box should statistically be encountered every 252 bp, therefore seven were expected in the 3 kb region upstream of *KFB^T*. However, a total of 184 were identified due to the excessive AT-richness of the promoter region. It is difficult to distinguish TATA boxes from insignificant strings of AT repeats; only four of the 184 candidates were bordered by cytosine or guanine bases.

The region upstream of *KFB^T* was also screened for initiator elements (1 in 128 chance: [CT][CT]A.[TA][CT][CT]), of which 43 were identified. It can function in place of a TATA box (Burke & Kadonaga, 1996), or the two may act in concert when present together (Smale & Kadonaga, 2003). There is usually an initiator element that overlays the transcription start site (Juven-Gershon & Kadonaga, 2010) and thus was here used as a marker to help locate the position at which *KFB^T* transcription commences.

Various attempts have been made to elucidate a transcription start site consensus sequence, such as: TCA[GT]T[CT] in *Drosophila* or [CT][CT]AN[TA][CT][CT] in mammals (Javahery *et al.*, 1994). The most reliable – albeit vague – of these is the YR rule found in 77 % of *Araibdopsis* promoters and also confirmed in rice (Yamamoto *et al.*, 2007). This stipulates that most transcription start sites are guanine or adenine and the immediately preceding base is cytosine or thymine. Patterns of up to 5 bp were queried but the authors could not reliably extend the rule beyond a single nucleotide.

One initiator element was found to be located 29 bp downstream of an 8 bp TATA box candidate that was situated within a GC rich region (Figure 3.3.5-3). This would have positioned the *KFB^T* transcription start site at 1644 bp upstream of the start codon. When compared against all other 5'-UTR lengths from the annotated *P. vulgaris* genome (Figure 3.6.1-2 upper), this figure was found to be in the top 1.15 %. The average 5'-UTR length in *P. vulgaris* was 191 bp. This was supported by studies in *Liliopsidae* in which the average 5'-UTR length was confirmed to be 129.8 bp (Pesole *et al.*, 2001).

A dot matrix alignment was carried out between the *KFB^T* nucleotide sequence and its own reverse complement to screen for the presence of self-complementary or repetitive regions (Figure 3.6.2-1). Matches between 7 and 15 bp were identified. Regions containing a sparse density of matches corresponded to areas of higher GC content, such as the *KFB^T* coding region. The manually predicted transcription start site was within an area of slightly increased GC content. However, raised GC content may also be indicative of coding

regions. A search was thus carried out for alternative open reading frames surrounding the *KFB^T* coding sequence.

A total of nine 5'-3' open reading frames exceeding 20 amino acids were identified upstream of the *KFB^T* start codon (Figure 3.6.2-3) and most corresponded to areas of higher GC content (Figure 3.6.2-1). Transcripts predicted from the RNA-Seq data were also mapped to the *KFB^T* sequence and largely conformed to two regions: the *KFB^T* transcript from approximately 900 bp upstream of the start codon and a second ~700 bp transcript beginning at position -2000 (Figure 3.6.3-1). The -900 bp region corresponds to areas of progressively increasing GC content toward the *KFB^T* start codon (Figure 3.6.2-1).

To further assess the authenticity of these transcripts, their expression levels were inferred by coverage depth in the RNA-seq dataset (Figure 3.6.3-3). The more distant transcript was significantly expressed in thrum, short homostyle, Somerset long homostyle and Chiltern long homostyle morphs – with greater expression than *KFB^T* in all but the latter sample. The three open reading frames contained within this region were used in a BLAST search that yielded no matches (not shown). This could represent a previously unidentified non-coding RNA within the *S* locus. The previously predicted transcription start site (1644 bp upstream of the start codon) is within this spuriously expressed region. This could explain its situation within a GC-rich region and consequential probable false identification as a transcription start site.

In conclusion, the true *KFB^T* transcription start site is likely represented by emerging RNA-Seq coverage approximately 900 bp upstream of the *KFB^T* start codon in all samples (Figure 3.6.3-3). However, coverage momentarily drops to zero in all three homostyle morphs approximately 300 bp into the *KFB^T* coding sequence before transcription appears to subsequently resume. This was only observed in one of the four thrum samples but in all twelve of the homostyle biological repeats. The short homostyle, Chiltern and Somerset long homostyle datasets overall contained an average of 17.76 %, 3.9 % and 2.7 % more reads than the thrum samples and therefore had greater chance of containing a read that spanned this gap. The contrary occurred instead and so the significance of this potential premature transcription termination is reinforced.

The homostyles present opposite mutant phenotypes but share one commonality in the loss of their self-incompatibility systems. The premature termination site directly

corresponds to the end of the first Kelch motif. Total loss of the remaining Kelch domain – especially the second blade vital for direct protein-protein interaction (Li *et al.*, 2004) – would result in dysfunctionality of target protein recruitment. It could be possible that *KFB^T* plays a role in the thrum self-incompatibility pathway but loss of *GLO^T* (short homostyle mutant) or *CYP^T* (long homostyle mutant) is indirectly associated with premature transcription termination of *KFB^T* or cleavage of its completed mRNA strand to disable the resultant protein.

3.9.6 Translational Regulation of *KFB^T*

There are four complete open reading frames between *KFB^T* and the predicted transcription start site approximately 900 bp upstream of the start codon (Figure 3.6.3-3). The scanning model of translation proposes that initiation occurs at the first AUG encountered by the 40S ribosomal subunit (Kozak, 1999). Although 15 to 50 % of genes are believed to possess upstream start codons (Mignone *et al.*, 2002), the number of alternative open reading frames prior to *KFB^T* presents a challenge in comprehending its translation. It has otherwise been suggested that, even if an upstream AUG initiates translation, the 40S ribosomal subunit may hold onto the mRNA after encountering a stop codon and resume scanning to reinitiate translation at a downstream AUG (Pesole *et al.*, 2001).

In this way, multiple proteins could be obtained from a single mRNA strand (Kozak, 1991), although translation efficiency would be expected to decline (Pesole *et al.*, 2001). However, *KFB^T* also contravenes this model because reinitiation becomes infeasible if the upstream open reading frame is greater than 30 amino acids long (Luukkonen *et al.*, 1995). It has further been speculated that leaky scanning may occur, in which upstream start codons situated within areas of low GC content are bypassed to favour those downstream (Davuluri *et al.*, 2000; Suzuki *et al.*, 2000) but the alternate open reading frames of *KFB^T* generally correspond to regions of increased GC content instead. Extended 5'-UTR lengths are known to improve translational efficiency and secondary structures downstream of the start codon can serve to reduce the risks of leaky scanning by slowing the ribosome to increase the chance of AUG recognition consequently initiate translation (Kozak, 1991).

Approximately 230 potential RNA hairpin structures were identified across *KFB^T* and its surrounding regions. Only perfectly self-complementary stems were screened for but

multiple bulges from non-matching bases can be structurally essential in functional RNA hairpins (Hsue & Masters, 1997). These stem-loop structures regulate translation by presenting protein binding sites and can also act in cellular localisation of RNA (Svoboda & Cara). The length of stem-loop components were limited to screen for the most frequently occurring hairpins (Schudoma *et al.*, 2010) but this could be expanded to include ~30 bp sites common for RNA binding proteins, such as iron response elements (Araujo *et al.*, 2012). A signature of loop regions within iron response element binding sites (CAG[TA]C[TCA]) was found in the distal non-coding RNA of *KFB^T* (not shown) but it was disregarded for not being bordered by a pair of self-complementary stem features.

One RNA hairpin was found to span the *KFB^T* start codon with a 5 bp stem and a loop length of 8 nucleotides. This could have served to inhibit *KFB^T* translation but the stem structures have a GC content of 0 % and a negative correlation has been observed between translational efficiency and the GC content of RNA hairpin stems (Babendure *et al.*, 2006). In addition, Kozak (1986) found that even a moderately stable hairpin across the start codon did not impair translation. It has been shown that secondary structures at the 5'-end of mRNA are most impenetrable by 40S ribosomes (Kozak, 1989; Kozak, 1994). Figure 3.6.3-3 suggests transcription begins ~900 bp upstream of the *KFB^T* start codon and there were fourteen potential stem-loop structures between position -800 and -1000 bp (Figure 3.6.2-3).

Translational efficiency can also be enhanced by secondary structures in the 3'-UTR too. The average 3'-UTR in *Liliopsidae* species was 273.3 bp long. These usually exceed 5'-UTR lengths (Pesole *et al.*, 2001) but this is not apparent in *KFB^T*. A transcription termination signal (AAUAAA) is situated 97 bp downstream of the stop codon and appears to correspond with the end of mRNA strands in Figure 3.6.3-3 (Connelly & Manley, 1988). A 16mer in the 3'-UTR of *Trypanosoma brucei* was found to be essential for translational efficiency of a procyclin protein (Hehl *et al.*, 1994). There were four RNA hairpins identified in the region downstream of *KFB^T*, one of which began at the central base of the stop codon. They all had 5 bp stems and loop lengths of 19, 15, 14 and 4 nucleotides.

The dot matrix alignment (Figure 3.6.2-1) identified two self-complementary matches between the 3' and 5' untranslated regions that would lead to circularisation of the transcript. It has been speculated that this could enhance translation (Gingras *et al.*, 1999) – although the model has no founding evidence. Unfortunately, both matches extend

beyond the 3'-end of the mRNA strand by at least 250 bp. Two perfectly palindromic 14 bp fragments were also identified in the 3'-UTR but it is improbable that they could fold over on themselves to form a structure without an intervening linker region.

3.9.7 Differential Expression of *KFB^T* and Other *P. vulgaris* Genes Between Morphs

Expression levels of *KFB^T* were quantified between flowers of pin, thrum, short homostyle, Chiltern long homostyle and Somerset long homostyle plants. If the *in silico* observations of premature transcription termination (Figure 3.6.3-3) are mirrored in nature then varying *KFB^T* transcript levels would have negligible efficacy in homostyle flowers.

In support of the current model regarding mediation of floral heteromorphy, no transcripts were observed in pin samples (Figures 3.6.3-1/2/3). The Chiltern long homostyle flowers presented similar *KFB^T* transcript levels to wildtype thrum, however *KFB^T* expression in the Somerset long homostyle samples appeared to be vastly increased (Figure 3.7.2). No phenotypic differences between the Chiltern and Somerset long homostyle mutant lines are currently known. Such spurious results must be rechecked via wet lab methods (Chapter 4).

A significant reduction in *KFB^T* expression was observed in the short homostyle flowers (Figure 3.7.2). The short homostyle phenotype is linked to a mutation in the *GLO^TS* locus gene, which is a MADS box transcription factor (Li *et al.*, 2016). If *GLO^T* activates *KFB^T* expression then this would infer that *KFB^T* is natively responsible for raised anthers in the wildtype thrum and its failed transcription leads to reduced anther height in short homostyle flowers. MADS box transcription factors (such as *GLO^T*) bind to CArG box promoter elements to activate transcription (Reichmann *et al.*, 1996). It follows that the presence of a sequence matching the CArG box consensus was confirmed 1865 bp upstream of the *KFB^T* start codon (Figure 3.6.1-1), ~900 bp upstream of the predicted transcription start site (Figure 3.6.3-3).

Genes demonstrating the most significant differential expression between *P. vulgaris* flower morphs were isolated (Figures 3.8-1/2/3). The Somerset and Chiltern samples were collectively treated as one long homostyle dataset. This enabled the three-directional comparisons between morphs (Figure 3.8-2) but analysis of *KFB^T* expression (Figure 3.7.2) suggests there may be significant genetic differences between these two mutant lines. As *CYP^T* is mutated in long homostyles (Huu *et al.*, 2016) and absent in pin (Li *et al.*, 2016),

differentially expressed genes common to both plants may possibly result from dysfunction of that gene. Genes only differentially expressed in long homostyle lines and not pin plants may highlight the effects of *CYP^T* mutation while in the presence of a functioning *S* locus remainder.

This same logic applies to hypotheses regarding short homostyle plants with respect to *GLO^T* mutation. Differentially expressed genes common to both short and long homostyles would be candidates for the self-incompatibility system, which is broken down in these mutants. There were 39 significantly differentially expressed genes common to all three morphs and these may demonstrate sensitivity of the *S* locus network and highlight the impact of modifying any one (or more) of its genetic components.

A probable carotenoid cleavage dioxygenase gene was found to exhibit minimal transcription in *P. vulgaris* long homostyle flowers and moderate expression across the other three (Figure 3.8-3). Differential expression of genes associated with carotenoid biosynthesis was also observed between compatible and incompatible pollination in *P. maximowiczii* (Lu *et al.*, 2018). Similar transcriptomic comparisons between developing pin, thrum and long homostyle styles of *P. oreodoxa* identified an average of 5 thousand differentially expressed genes (Zhao *et al.*, 2019). Approximately 1.2 thousand genes were found to be differentially expressed during self-incompatibility of *P. maximowiczii* (Lu *et al.*, 2018).

This *KFB^T* analysis produced a list of the most differentially expressed genes between the *P. vulgaris* morphs. These were used to query the BLAST database. No matches to *KFB^T* were found. This was expected because *KFB^T* only represents between 0.0005 and 0.002 % of the *P. vulgaris* transcriptome (Figure 3.7.2) and the filter thresholds were arbitrarily set much higher (Figure 3.8-1). Values for the genes presented by this analysis were up to almost 5000 transcripts per million and differential expression of *KFB^T* was thus much too subtle for detection via this method. It is essential for the Gilmartin lab to pursue this list of genes and accomplish characterisation of the end-to-end pathways utilised by floral heteromorphy. This dataset was readied for differential expression analysis of candidate genes encoding potential *KFB^T* partner proteins involved in floral heteromorphy (Chapter 6).

Chapter 4

4 Spatial and Temporal Expression of *KFB^T*

As the sequence of events leading to floral heteromorphy during bud growth has been defined (Webster & Gilmartin, 2006), identifying the developmental phase that corresponds with *KFB^T* expression would offer potential insight into its function. Transcript levels of *KFB^T* had previously been quantified by RNA-Seq in flowers of thrum, pin, short homostyle and long homostyle plants (Chapter 3.7). The subsequent objective was therefore to analyse how the quantity of *KFB^T* transcripts changes throughout development of the flower bud.

It was also hypothesised that gaining information about where *KFB^T* expression is localised would give insight into which part of the flower the encoded protein played a role in the control of heterostyly. This could be combined with knowledge about which aspects of heteromorphy are affiliated with that floral region and therefore potentially provide early indication of *KFB^T* function. Further analyses were therefore carried out to locate the floral whorls in which *KFB^T* is expressed. A qPCR approach was used for these investigations (Chapters 2.8.3-2.8.5).

Aside from analysing the transcriptional activity of *KFB^T*, polyclonal antibodies were evaluated to observe the timing and location of *in situ* *KFB^T* protein production (Chapter 2.19). These assays were designed to either obtain supporting data that reinforced the qPCR results or to highlight variance between transcript levels and post-translational quantities of *KFB^T* that would demonstrate regulatory networks or protein accumulation, aiding to establish a model for *KFB^T* activity and its function in floral heteromorphy. Further to these immunolocalisation experiments, a vector containing the *GUS* reporter gene driven by the *KFB^T* promoter was also constructed (Chapter 2.11.6) for *in planta* observation of *KFB^T* in the flower (Chapter 4.8).

A fluorescent dye is used in qPCR that binds to DNA (Thermo Fisher, 2016). Therefore, fluorescence increases as DNA is duplicated during each cycle of the polymerase chain reaction while in linear phase (Figure 4). This is monitored in real time by the qPCR thermocycler machine. The fluorescence at each cycle is plotted on a graph and forms a characteristic qPCR curve (Figure 4). This consists of an exponential amplification phase followed by a linear phase, in which template quantity is doubled after every cycle (Yuan *et al.*, 2006). The PCR reaches a final plateau phase when the reaction is saturated because

all free primers have been incorporated into products and amplification consequently ceases (Yuan *et al.*, 2006).

The user sets an appropriate fluorescence threshold that intersects all sample graphs at their linear phase of amplification (Thermo Fisher, 2016). The cycle number at which a sample meets this fluorescence threshold is called a C_q (quantitation cycle) value and these figures are compared between samples. The fluorescence at this point is relative to the amount of target template present at the beginning of the qPCR, therefore using cDNA (which is made from RNA transcripts) as template provides indication of the target gene expression levels in the original sample. For example, if Sample A reaches the fluorescence threshold at cycle 35 and Sample B meets it at the 36th cycle, it can be concluded that sample A contains twice the amount of target template than sample B, thus the target gene was expressed doubly in Sample A. If the living tissue was immediately frozen and handled properly, qPCR produces a snapshot of gene activity for the organism at that moment of sampling. This method was therefore used to compare gene expression levels of *KFB^T* in different floral whorls and to observe how this changes throughout flower bud development.

Real-Time Quantification of PCR Products

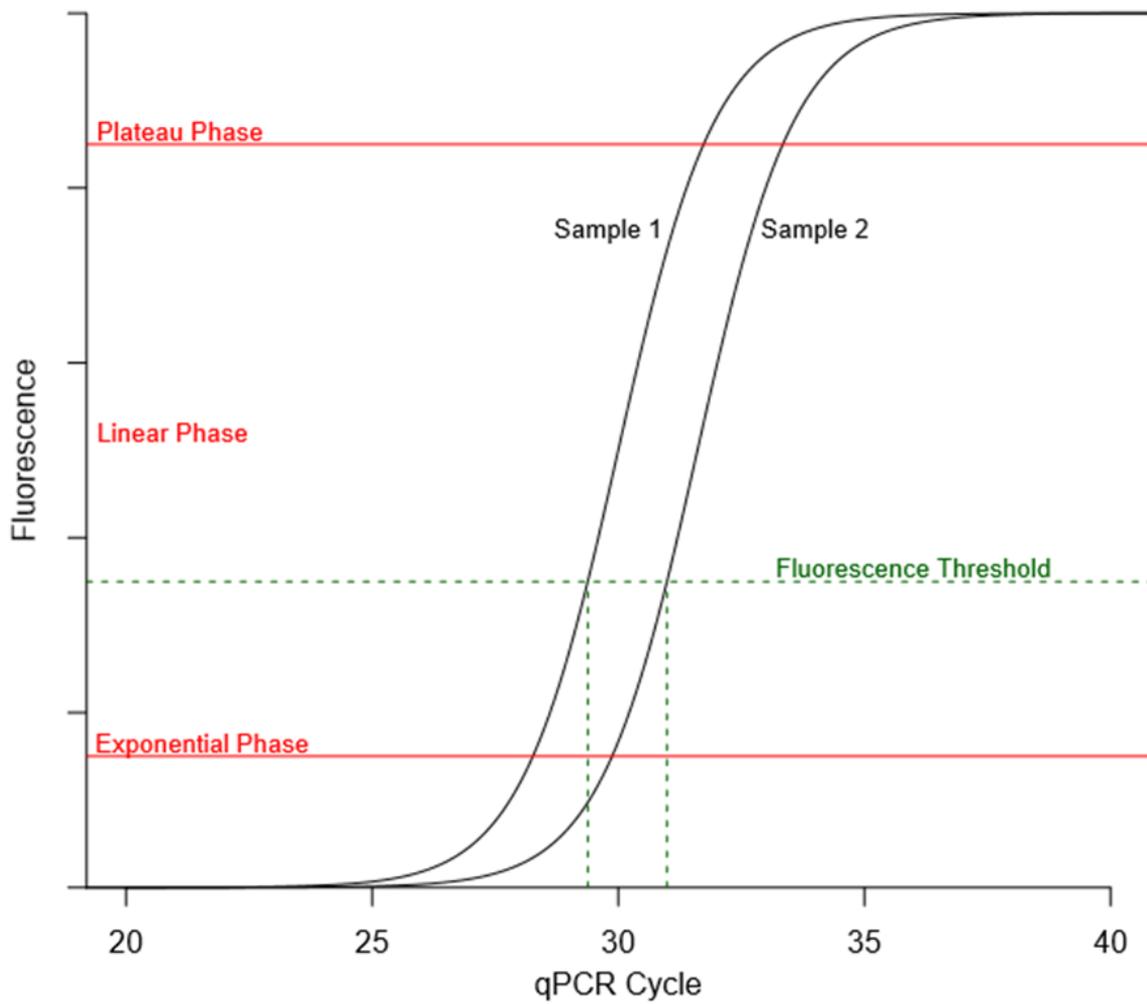


Figure 4: Exemplar qPCR fluorescence graph depicting the three phases of product amplification and the fluorescence threshold used to compare template quantity between samples.

4.1 Efficiency and Design of *KFB^T* qPCR Primers

Experiments were designed to quantify *KFB^T* transcript levels throughout the flower during development, to elucidate the location and timing of its activity. The efficiency of two primer pairs designed to the *KFB^T* coding sequence was calculated and tested for qPCR suitability (Chapter 2.8.4). A pair of previously designed reference gene primers were also included as a positive control (Kent, 2016). These *PP2A* primers amplified with an efficiency of 99.71 % (Table 4.1). This showed the test functioned correctly and reconfirmed validity of the primers.

Table 4.1: Efficiency of two *KFB^T* qPCR primer pairs and those for a *PP2A* reference gene used as a positive control.

Primer Pair	Efficiency /%
<i>KFB^T</i> #1	112.00
<i>KFB^T</i> #2	80.59
<i>PP2A</i>	99.71

The two *KFB^T* primer pairs performed with 112 % and 80.59 % efficiency. The ideal range lies within 90-110 % efficiency (Thermo Fisher, 2016) and so the former primer pair was used (F: TGATTGGGACGGGATGAGT; R: CTA CTGGTGTCTATCCGCT). Efficiency exceeding 100 % is caused by template impurities that exhibit decreasing inhibition on the reaction as the template is diluted (QIAGEN, 2010). The selected forward primer began at position 560 of the *KFB^T* nucleotide sequence and so was situated sufficiently downstream of most cleaved regions (Figure 3.6.3-3) identified in homostyle transcripts, which had an average end site at position 492 (Figure 4.5-1).

It would be invalid to compare the expression levels of two genes using primers with different efficiencies. Although efficiency of the primer pairs were checked here, this series of qPCRs only screened the *KFB^T* gene and all assays used the same primer pair, therefore any errors would have been systematic across all samples and would not have affected relative expression patterns.

4.2 Temporal Expression of *KFB^T* in *P. vulgaris* Thrum Flower Buds

To elucidate how *KFB^T* expression levels change throughout flower development, qPCR was carried out on thrum flowers of four different stages. This indicated in which aspect of floral heteromorphy the gene acts and also served to highlight the stage of flowering that other experiments should focus on. Floral developmental stages were defined in Chapter 2.1.1. A second set of flowers was sampled as a biological repeat. The qPCR was carried out (Chapter 2.8.5) using the *PP2A* reference gene (Chapter 2.8.3).

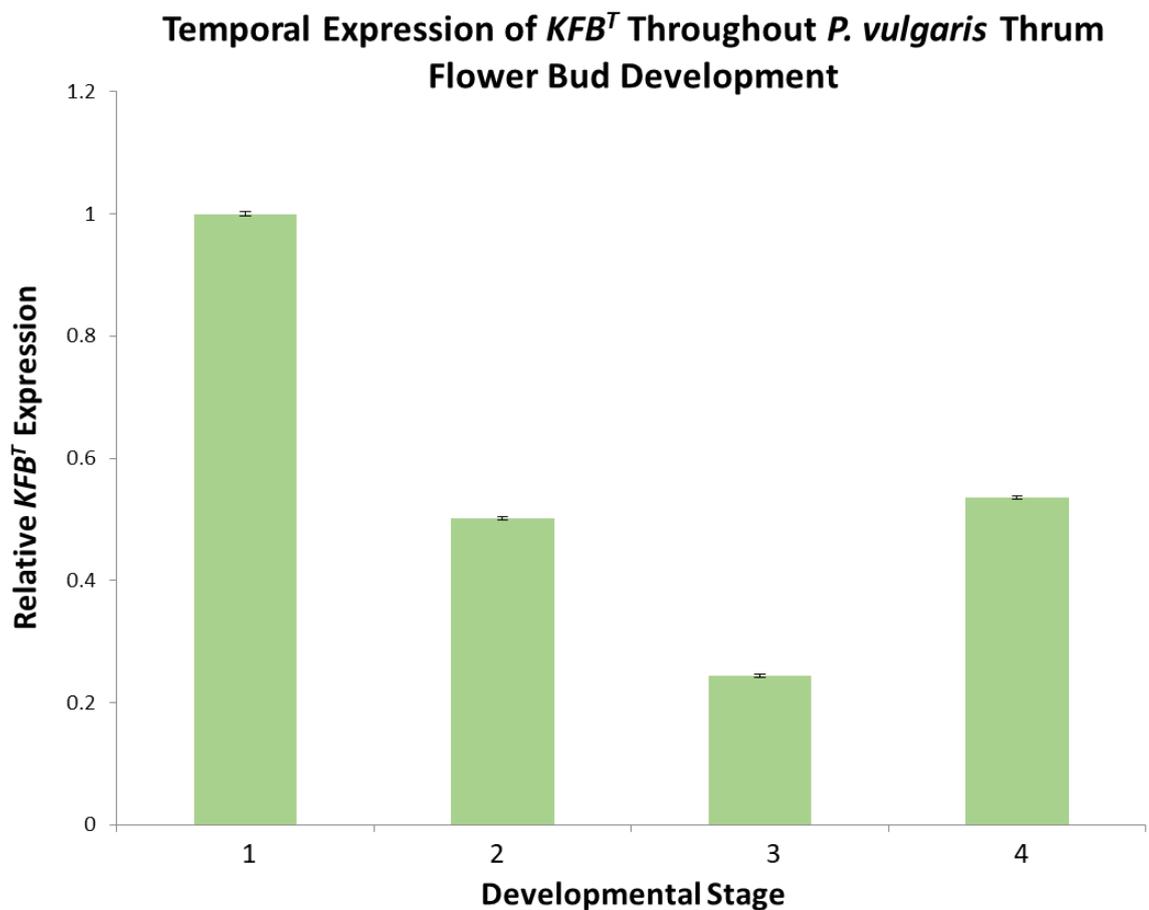


Figure 4.2: Quantification of *KFB^T* transcripts across four developmental stages of thrum flowers, from ~5 mm buds to fully open mature flowers. Data is normalised to stage one and was carried out across two biological repeats against a *PP2A* reference gene. Standard error bars are shown.

The greatest *KFB^T* transcriptional activity appeared in the youngest samples at stage one of development (Figure 4.2), approximately ~5 mm in length. The earliest anatomical aspect of heteromorphy can be observed at this stage and occurs as the suppression of style length in thrum buds compared to pin (Webster & Gilmartin, 2006). Expression subsequently decreased in stage two and declined further at stage three before rising to the second highest overall levels in stage four. This increase upon opening of mature flowers was consistent across all repeats and signifies a clear resurgence of *KFB^T* expression in the latter phase of flower development. Heteromorphic features linked to this latter stage include pollen size, shape and potential preparation of the self-incompatibility system.

4.3 Differential Spatial Expression of *KFB^T* Between *P. vulgaris* Morphs

Transcript levels of *KFB^T* were quantified in the dissected flowers of thrum, short homostyle and long homostyle buds. Elucidating how *KFB^T* expression levels vary throughout the flower itself allowed identification of where the gene acts. This offered insight into *KFB^T* function by relating the location of its expression to the known aspects of floral heteromorphy affiliated with that location. Inclusion of the homostyles allowed further functional deductions to be made corresponding between the mutant phenotypes and expression anomalies in that morph. This also highlighted links at the genetic level, by observing how mutations in *GLO^T* or *CYP^T* effect *KFB^T* expression in short and long homostyles, respectively (Li *et al.*, 2016; Huu *et al.*, 2016).

A pool of thrum buds containing mixed developmental stages were dissected into their four floral whorls for this experiment. Aside from the technical triplicate repeats for each sample, the qPCR was repeated three times to include assays with the *alpha tubulin*, *ELF1 α* and *PP2A* reference genes (Chapters 2.8.3 & 2.8.5). This improved reliability if any of the reference genes were not expressed with perfect stability in these different tissue types.

The *ELF1 α* reference gene was also used to screen the dissected whorls of a *P. vulgaris* long homostyle mutant. The long homostyle phenotype is caused by a mutation in the *CYP^T* *S* locus gene that leads to failed suppression of style length (Li *et al.*, 2016; Huu *et al.*, 2016), so this assay highlighted any *KFB^T*-*CYP^T* relationship and looked for a possible *KFB^T* role in the style. The short homostyle mutant was also screened in the same way. This phenotype is caused by a mutation in the *GLO^T* *S* locus gene that leads to anthers situated lower in the corolla tube (Li *et al.*, 2016).

Spatial Expression of *KFB^T* In Various *P. vulgaris* Floral Morphs

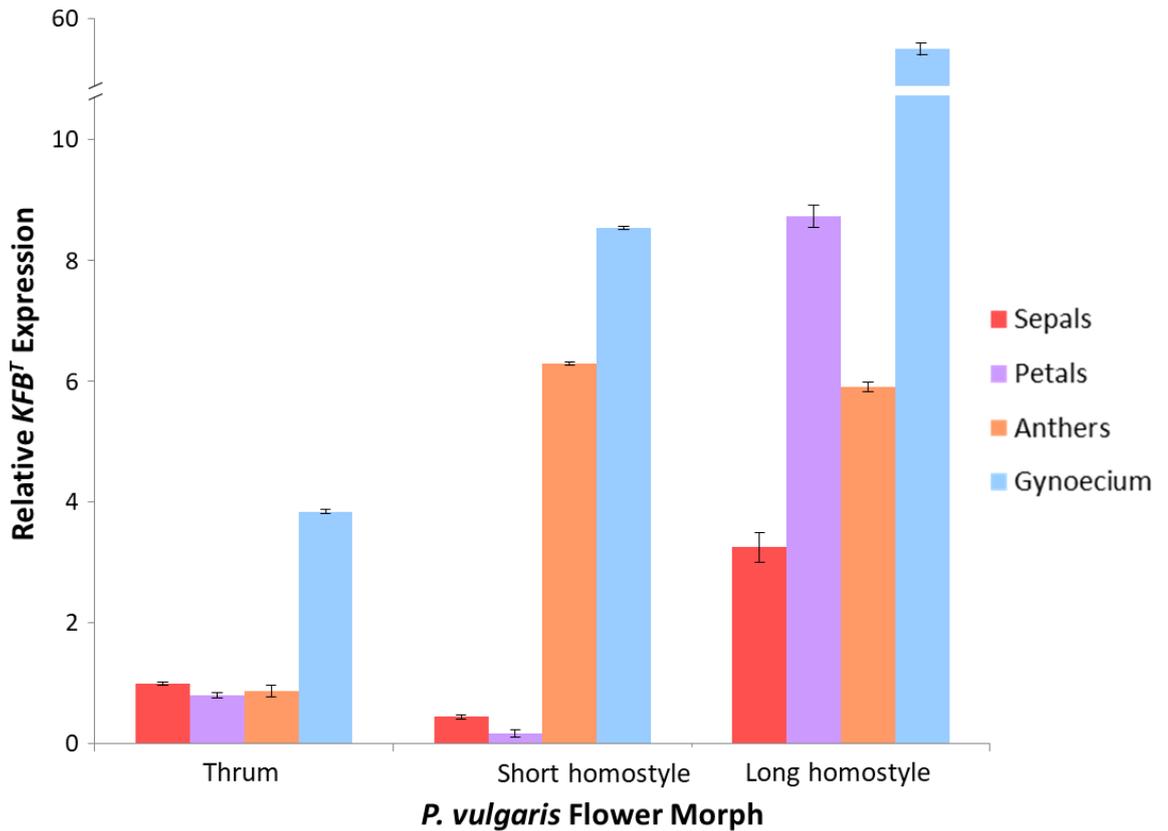


Figure 4.3: Quantification of *KFB^T* transcripts throughout four floral whorls in thrum, short homostyle and long homostyle flowers. Gynoecia transcription of *KFB^T* was significantly higher in long homostyles. Expression was almost entirely lost in short homostyle petals. Data is normalised to thrum sepals. Standard error bars are shown.

Highest levels of *KFB^T* transcripts were consistently observed in the gynoecium, which consisted of the stigma, style and ovary (Figure 4.3). This is the site of style length dimorphism; the most obvious anatomical aspect of floral heteromorphy. The gynoecium is a site of action for self-incompatibility. Although the utilised primer pair was designed downstream of the region potentially cleaved from the centre of most homostyle transcripts (Figure 3.6.3-3), some of the missing regions varied greatly from this average size and position. However, these qPCR samples appeared to be unaffected by this and long homostyles actually exhibited the greatest overall quantity of *KFB^T* transcripts, especially in the gynoecium, which here far exceeded levels of all other samples. The gynoecium in this morph is effected by mutation of the *CYP^T S* locus gene and presents a long pin-like style (Li *et al.*, 2016; Huu *et al.*, 2016).

Short homostyle samples demonstrated almost a total loss of *KFB^T* expression in petal tissue. As the anther filament is fused within the petal, this whorl is responsible for anther height dimorphism. Shortening of the anther filament in this morph has previously been linked to a mutation in *GLO^T* (Li *et al.*, 2016). The short homostyles also appeared to exhibit

reduced KFB^T levels in the sepals, however the only mutant line available at this time were of the *Hose-In-Hose* variety in which the sepals are biologically converted to a whorl of petals (Webster & Gilmartin, 2003). It can therefore be concluded that the reduction of KFB^T here was a mirror of that observed in the petals and not a true measurement of sepal expression.

The two homostyle mutants displayed raised KFB^T transcript levels in their anthers (Figure 4.3). Long and short homostyles both exhibit a breakdown of the self-incompatibility system and are therefore self-fertile (Dowrick, 1955; Ernst, 1955). These anther measurements include pollen, which is a possible mediator of self-incompatibility in *Primula*. This suggests a potential self-incompatibility role for KFB^T .

4.4 Relative Expression of *KFB^T* Throughout the Flower Across Development

A final effort was made to carry out a larger qPCR that repeated the spatial expression assay but did not mix the developmental stages. This consisted of sixteen samples comprising of four dissected floral whorls from four developmental stages of flower buds. Maintaining both dimensions allowed for detailed monitoring of how the location and quantity of *KFB^T* transcription changed over time throughout the flower. Obtaining this expression data across the life of the flower bud provided a high-resolution map of *KFB^T* activity. The PP2A reference gene (Chapter 2.8.3) was used in the qPCR (Chapter 2.8.5) and this was completed twice to obtain a technical replicate. A third attempt was made with a different sample set as a biological repeat.

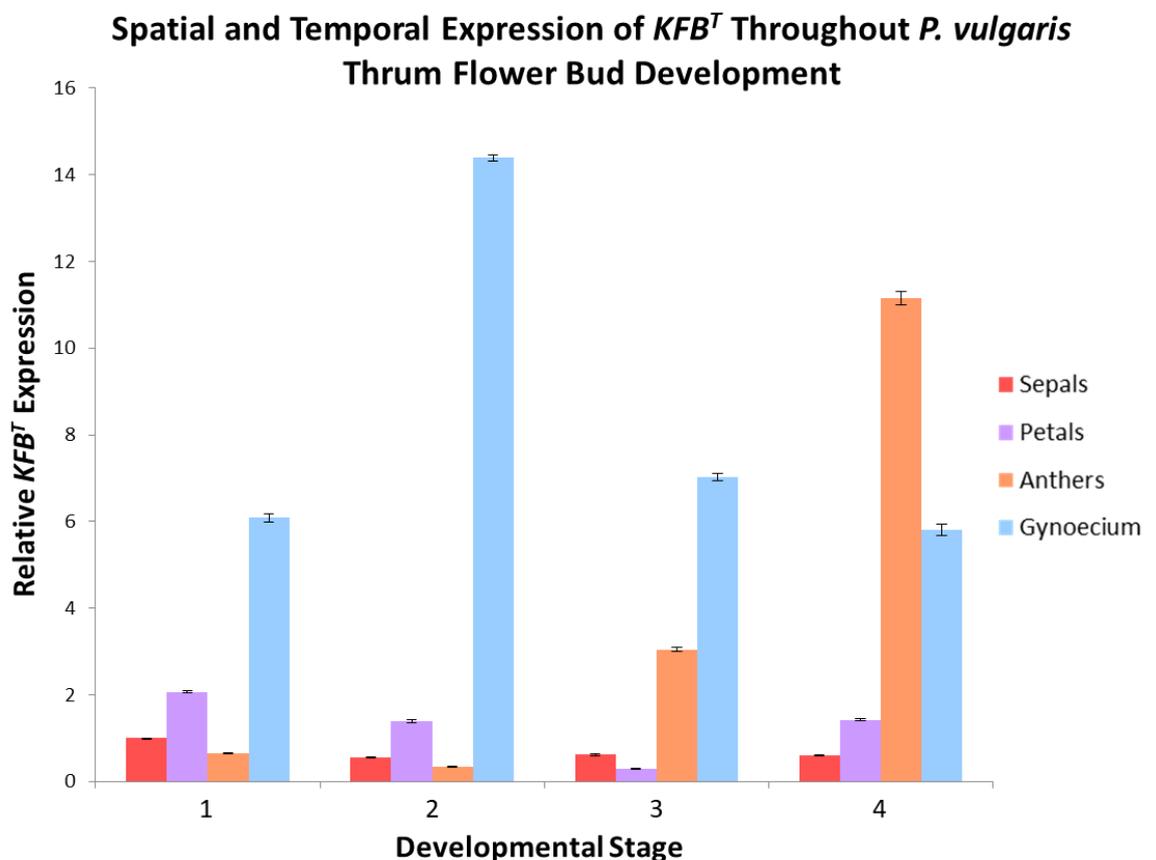


Figure 4.4: Quantification of *KFB^T* transcripts across the dissected floral whorls of thrum flower buds across four stages of development. Gynoecium expression of *KFB^T* was consistently high and transcription in the anthers increased toward flower maturity. Data is normalised to thrum sepals at stage one with the *PP2A* reference gene used as a control. Standard error bars are shown.

Expression of *KFB^T* in the sepals remained minimal and stable across flower development (Figure 4.4). There are no phenotypic aspects of floral heteromorphy associated with the sepals of thrum plants. Transcript levels in the petals presented a similar pattern to the average temporal expression data in Figure 4.2. A reduction of *KFB^T* transcription in the petals occurred at stage three of flower bud maturity. It is approximately at this stage when

cell shape is suppressed in the upper corolla tube and they remain globular while those in the lower corolla tube become cylindrical (Webster & Gilmartin, 2006).

Expression of *KFB^T* increased significantly in the anthers toward the latter two stages of development. Pollen maturation occurs during these phases and begins to shed from the anthers. It is therefore possible that *KFB^T* has a role in the pollen, which may mediate self-incompatibility.

Levels of *KFB^T* transcripts in the gynoecium were consistently high throughout development; the highest of all four whorls in the first three stages and superseded only by anthers upon opening of the mature flower. The greatest *KFB^T* expression was measured in the gynoecium of stage two buds and the only distinguishable feature of floral heteromorphy at this age is suppression of style length in thrums compared to pin flowers (Webster & Gilmartin, 2006).

4.5 Screening for Presence of *KFB^T* in Flowers of Four Developmental Stages

Prior to qPCR experiments, a non-quantitative PCR was used to initially screen for the presence of *KFB^T* in *P. vulgaris* flowers across four developmental stages. This presented a broad overview of temporal *KFB^T* expression to inform which developmental stages should be included in the downstream qPCR assays for observing more intricate details of dynamic *KFB^T* activity.

The PCR primers spanned a site (Figure 4.5-1) that was later identified via bioinformatical RNA-Seq analysis to contain a gap central to most homostyle transcripts (Figure 3.6.3-3). The use of long homostyles therefore also allowed assessment of this gap. It was important to detect intact transcripts and confirm that *KFB^T* does not entirely undergo premature transcription termination or post-transcriptional cleavage, which would render it non-functional in homostyle plants – regardless of any expression activity. The exact size and border positions of the cleaved region varied between samples but the reverse primer (Figure 4.5-1) was either within or after the missing portion in all available homostyle transcriptome datasets (Chapter 3.6).

```

ATGGAAGTTATTCCTGGTCTGCCTGAAGATCTTGGACTCGAGTGTATGATTCGTTTCACAT 60
TACACAACATTCCGAGTCGTTTCCCAAACATGCCATTTATGGAGAAAACCTTCTCCAAACC 120
ACAGATTTTCATAGTTACAGAAAGAACAAGGATACAGCCACAAAATGATTTGTTTTGTA 180
CAATCTATTCGCCAAATGCTCTTGCTGATGAAACTGGAAAATCAGCCAACTCGTGCCGT 240
TATGGGATCACTGTATTCGACGAAGTCGGACATGGGGGAGACTTAGTCAGGTTCCCAAGT 300
ACCAAAGCGGGCTCCCTTGTTTTGCCGTGTGGCGAGCTCCGGAAATAAGCTCATCGTTA 360
TGGGCGGGTGGGACCCGTTTAGTTACCACCCAGTTAAGGACGTGTTTGTTTATGATTTTG 420
TGAACCAGTTGTGGCGACAAAGTAAAGATATGCCGTCCAAACGATCATTTTTTCGCTATGG 480
GCGCCATAGATGTTAAGGCATGTATACGTCGCGGGAGGCCACGATGAGAATAAAGGTGCT 540
TTGAAATCAGCTTGGGTTTATGATTTGGGACGGGATGAGTGGACTGAGATGATACAGATG 600
GCACACGAACGTGACGAGTGTGAGGGAATAGTGATGGGTAATGAGTTTTGGGTAGTTAGC 660
GGATACGACACCAGTAGTCAAGGGGTGTTTGTGACGAGCGCAGAGTCTATTGTGTCAGT 720
ACCAGGATGTGGAATCTAGTTGAATCTGTATGGAAGGCGGGCCAGTGCCCGAGGTCATCT 780
GTCCTAAGTTTAAAACCTAGTCAGTTGATAAGTTATAACGAATTCAGCTCGGCTATTACA 840
GATGGAGCATTGTTGGGATCGCGCTGGGCGCCAGATTCTTTAAAGGAGTCAGCAGATGTT 900
GATGTAAAGAAAGCTTTTTTTTTAGTGGATGTTGGTGAAGGGCAAACTATAGAATCGAG 960
AAGATTAATGTGCCTGATCAGTTTTCTGGTTTAGTTCAATCTGGCTGTCTGTTGAAAT 1020
TGA
Start Codon Stop Codon Forward Primer Reverse Primer Missing Region

```

Figure 4.5-1: Positions of primers designed for non-quantitative amplification of *KFB^T* from flower cDNA samples. Primers spanned a region that was earlier identified by RNA-Seq to be commonly absent in homostyle transcripts. The average size and position of this missing region is shown.

Expression of *KFB^T* was identified in all developmental stages of long homostyle samples from the Somerset population (Figure 4.5-2), which is indicative of continuous *KFB^T* activity throughout flower maturation. Transcripts were also detected in all stages of Chiltern long homostyle flower buds, though more variation was observed than in Somerset samples. The actin positive control bands showed consistent integrity of the cDNA pools, proving this variance was not due to a PCR flaw. Expression appeared extremely minimal at stage two of the first biological repeat, however the negative control lanes are clear, thus the faint band must represent true *KFB^T* transcription and not background contamination.

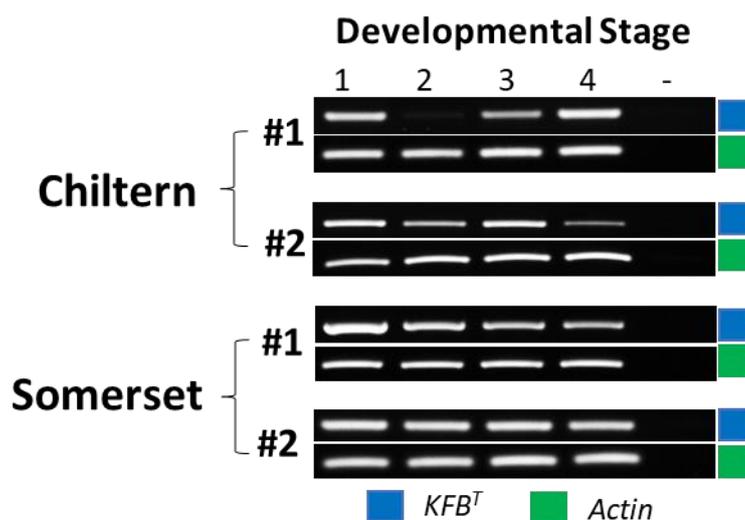


Figure 4.5-2: Non-quantitative detection of *KFB^T* expression across flower development. Intact transcripts were detected throughout bud maturation. Biological repeats were carried out on long homostyle lines from two UK populations. Four developmental stages were assessed, ranging from ~5 mm buds to fully open mature flowers. The actin gene was amplified as a positive control to prove consistent cDNA integrity across samples.

A gap central to most *KFB^T* RNA strands of homostyle flowers had been identified via RNA-Seq (Figure 3.6.3-3). This PCR assay importantly confirmed that all flower buds contain a detectable level of intact transcripts and therefore *KFB^T* transcription is not entirely prematurely terminated or post-transcriptionally cleaved, which would render it non-functional in homostyle plants. However, it remained unclear whether the variation observed throughout Chiltern flower bud maturation was representative of differential expression or the post-transcriptional modification identified from RNA-Seq (Figure 3.6.3-3).

This preliminary screen gave no conclusive indication of isolated stages at which *KFB^T* expression is categorically turned on or off, which would have provided strong functional evidence and a clear area for further scrutiny. No phenotypic differences between the

Chiltern and Somerset long homostyle lines have been recognised that would explain the observed variation. It was decided to include all samples in downstream qPCR experiments with a different primer pair to focus exclusively on the intricate mapping of *KFB^T* expression patterns.

4.6 Identification of a Suitable *KFB^T* Antibody

A dot blot was carried out (Chapter 2.19.2) to test interaction of the *KFB^T* antibody with the peptide that it was raised against. Antibodies from two rabbits were supplied as conjugates to keyhole limpet hemocyanin and bovine serum albumin proteins. After development of the film, no light emission had been captured at all (not shown). There was no sign of protein interaction with the antibody. As this could have been caused by a flaw in the dot blot procedure, the test was repeated via Western Blot (Chapter 2.19.4).

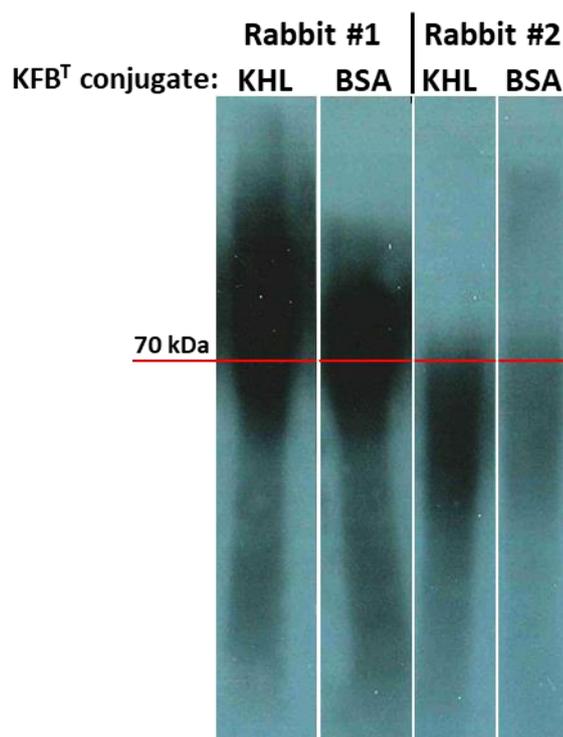


Figure 4.6-1: Affinity of antibodies raised in two rabbits against synthetic *KFB^T* peptide conjugates. The Western blot was carried out on a single sheet; white lines indicate where the photograph has been cropped to remove intervening blank lanes.

The antibody from rabbit one demonstrated affinity to the *KFB^T*-KHL and *KFB^T*-BSA conjugates (Figure 4.6-1). While rabbit two also displayed binding with *KFB^T*-KHL, it had far weaker affinity for the *KFB^T*-BSA conjugate. This suggested that the former antibody bound to *KFB^T*, whereas the second bound to the KHL tag instead. Both proteins presented a

smear instead of a neat band, however this was still deemed to represent affinity and so it was concluded that the dot blot method had failed but the Western Blot had confirmed the rabbit one antibody should be taken forward.

This antibody was subsequently tested against crude protein extract from *P. vulgaris* buds of mixed stages. Eight separate fractions were tested from the antibody purification step carried out by Dundee Cell Products. This tested for antibody interaction with the true KFB^T protein from *P. vulgaris* and not only the synthetic peptide conjugates they were raised against.

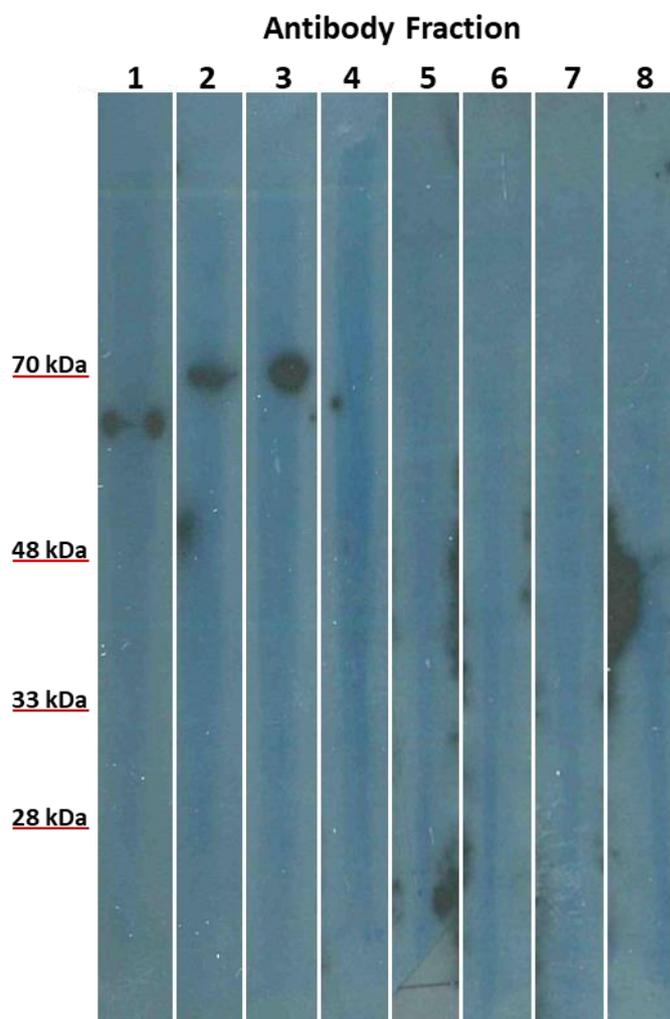


Figure 4.6-2: Antibody affinity for KFB^T from the crude protein extraction of ~5 mm *P. vulgaris* thrum flower buds. Eight fractions from protein purification were tested. The assay was carried out on a single sheet; white lines indicate where the photograph has been cropped to remove intervening blank lanes.

The second and third antibody fractions presented a distinct band on the Western Blot (Figure 4.6-2), however they indicated a 70 kDa protein but KFB^T is 38 kDa. Fractions five and eight presented smears that spanned the 38 kDa region but these were background blemishes off-centre from the protein lane. No discrete bands of the correct size were detected and the blotches are assumed to be background noise. Therefore, as this Western

Blot had failed to identify an antibody fraction with suitable functionality, it was decided to move forward with the most concentrated fraction in a final Western Blot attempt to investigate spatial and temporal production of the KFB^T protein in thrum flowers. Information provided by Dundee Cell Products showed that fraction one was the most concentrated and this appeared to perform better in initial tests (Figure 4.6-1).

4.7 Investigating Spatial and Temporal KFB^T Production via Western Blot

Flower buds from different developmental stages were used in an effort to gain temporal information regarding production of the KFB^T protein and spatial data was sought by dissecting samples into their four floral whorls. This was designed to obtain protein-level data in support of the transcript-level qPCR experiments.

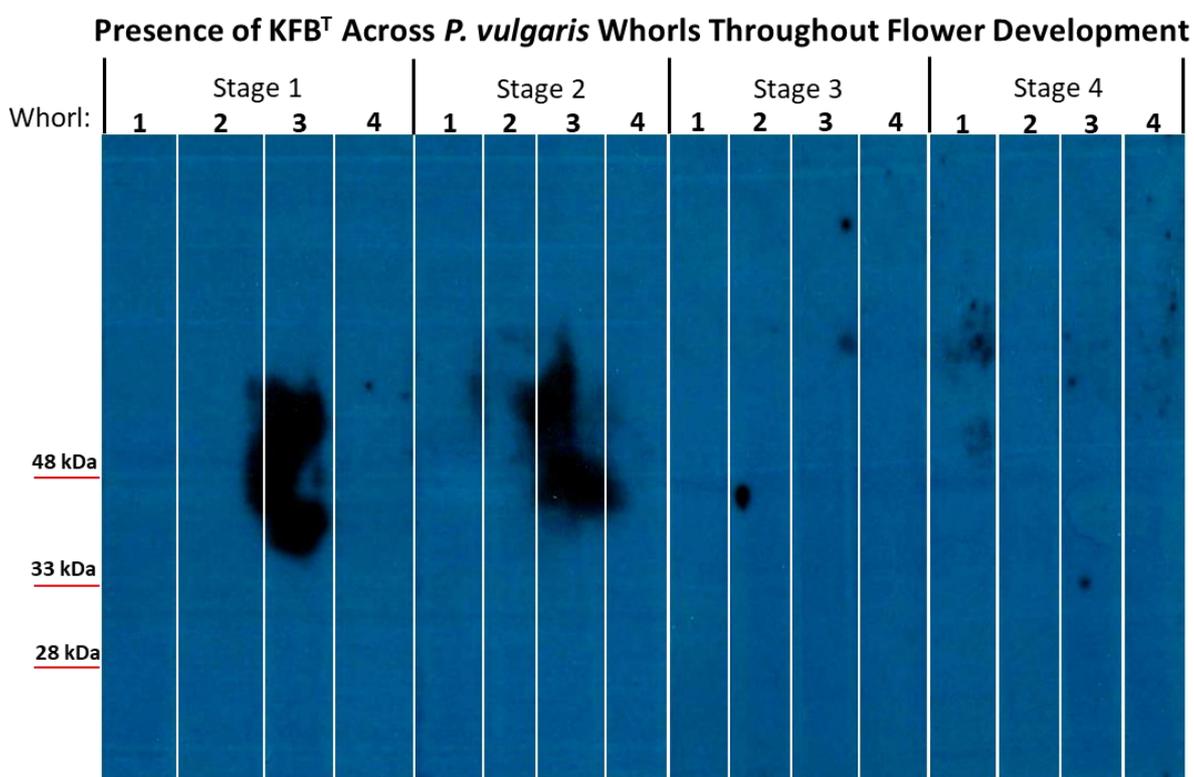


Figure 4.7: Antibody affinity for KFB^T in dissected whorls of *P. vulgaris* flowers throughout development. White lines are a visual aid to align sample lanes.

Two large smears appeared corresponding to the anther samples at the earlier two phases of maturation (Figure 4.7). This is contrary to qPCR data that indicated an increase of KFB^T transcription in the latter two developmental stages (Figure 4.4). The blotch at stage one crossed into the adjacent lane but did span the 38 kDa size marker where KFB^T is expected, however the second stage smear was above this size. It is unclear what is represented here

on the X-ray film but no reliable insights could be gained from these assays. It was concluded that the antibodies were incapable of eliciting the required function. The qPCR transcription data would instead have to be supported at the protein level via transformation of the *GUS* reporter gene construct driven by the *KFB^T* promoter (Chapter 2.11.6).

4.8 Visualisation of *KFB^T* Via *GUS* Reporter Gene Assay

The qPCR assays had identified the floral whorls in which *KFB^T* is transcribed, however a β -glucuronidase (*GUS*) reporter gene construct was made to complement these transcript-level studies and explore the cellular location of *KFB^T* expression in those identified whorls. A vector had been assembled so that, after transformation into the target organism, the *KFB^T* promoter would be activated *in vivo* but had been modified to drive *GUS* expression instead (Figure 4.8-1).

The technique works by taking samples from the transformed organism and incubating them with glucuronide substrates that interact with the GUS enzyme to turn it blue (Jefferson *et al.*, 1987). In this way, the exact regions of *KFB^T* production could be visualised. For the purpose of gaining data efficiently, the vector was transformed into *Arabidopsis*. This also served to elucidate whether *KFB^T* is regulated by a fundamental housekeeping gene or requires a more *Primula*-specific transcription factor.

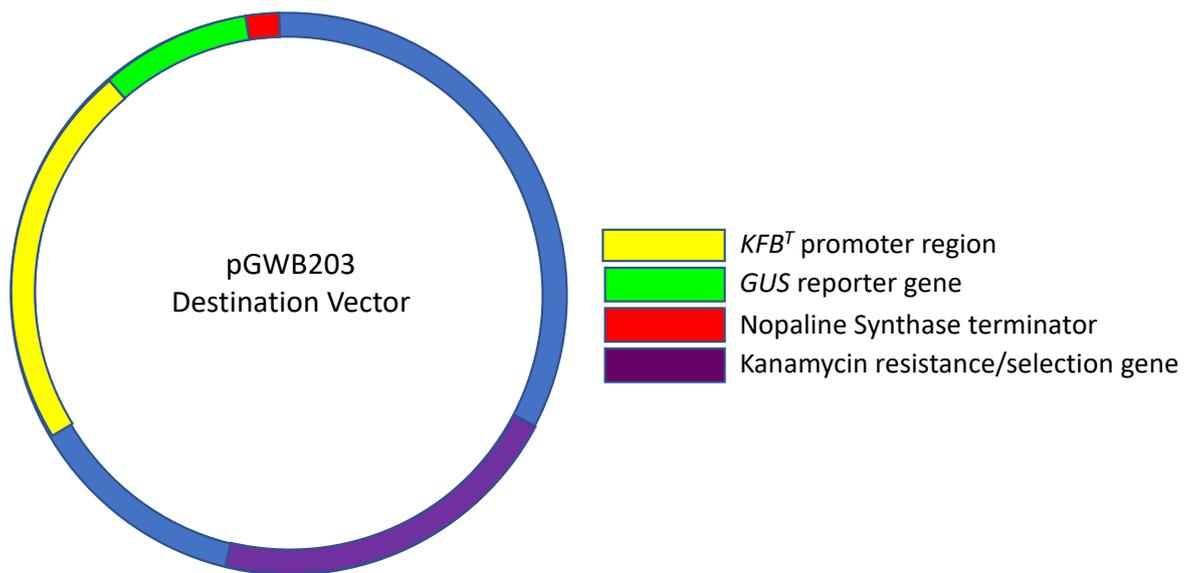


Figure 4.8-1: Main features of the construct in which the *KFB^T* promoter region drives expression of the *GUS* reporter gene to visualise location of *KFB^T* protein production.

After establishing transgenic *Arabidopsis* lines (Chapter 2.16.1), one leaf sample presented blotches of GUS activity (Figure 4.8-2B) that followed no patterns of venation or stomatal cells. It was observed in no other leaf samples and is believed to be spurious misexpression of GUS, likely caused by bacteria. Blue GUS staining was also noted on a leaf that appeared to strongly follow a tear wound caused by handling with forceps (Figure 4.8-2A). To investigate whether this signified stress-induced *KFB^T* transcription, leaves were laterally cut in half and added to GUS assay solution. The remaining half – still attached to the plant – was severed 90 mins later and also added to GUS solution. However, the results in Figure 4.8-2A were not reproducible.

Figure 4.8-2C depicts two very late stage flowers undergoing abscission after seed set and shedding of the siliques. They each had a single seed affixed to the remainder of their senescing siliques that appeared to have imbibed water during the GUS staining process and commenced germination. Both samples presented GUS staining in their carpels.

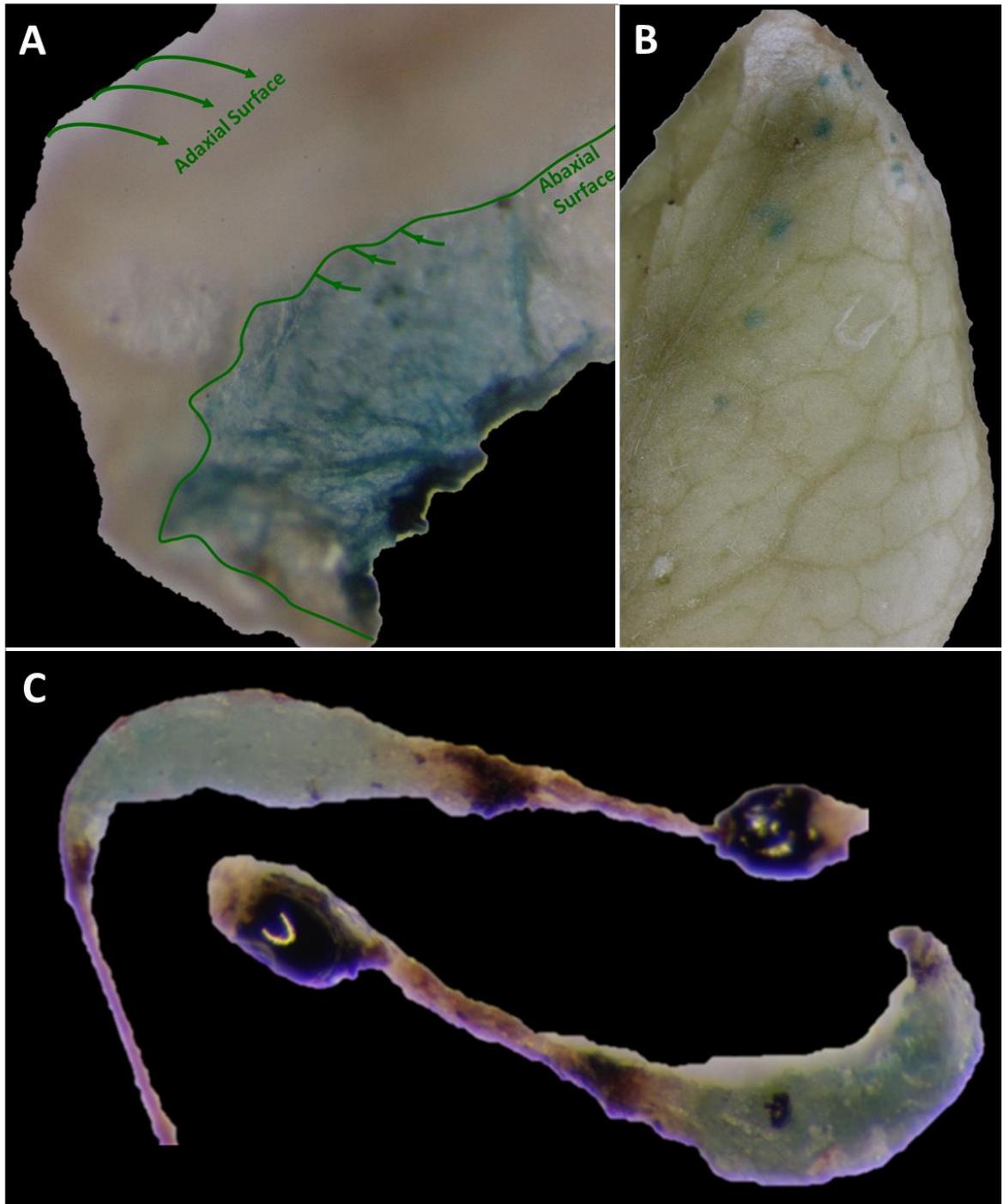


Figure 4.8-2: Location of GUS staining in transgenic *Arabidopsis* tissue when regulated by the *KFB^T* promoter and 5'-UTR regions. A) One leaf showed GUS staining around a wound in the torn tissue. The adaxial surface is out of focus beyond the depth of field because the sample had curled over due to dehydration under the microscope. B) Blotches of GUS activity were found on one leaf sample and followed no pattern of venation or stomatal cells. C) Two mature flowers (both pictured) undergoing post-seed set abscission presented GUS activity in the gynoecia.

4.9 Discussion

The temporal and spatial expression patterns of *KFB^T* have been investigated using qPCR on dissected floral whorls throughout development of thrum, pin and homostyle flower buds. This has been supported at the protein level through transformation of a vector containing the *KFB^T* promoter and 5' untranslated regions driving transcription of a *GUS* reporter gene in *Arabidopsis*. The GUS assay presented reporter staining in a torn leaf sample (Figure 4.8-2A). This finding was irreproducible and could have been caused by bacterial infection via the wound. Although it is widely accepted that endogenous GUS activity is absent in plants, endosymbiont bacteria have been known to display false positives in this way (Tör *et al.*, 1992). This is also believed to be the cause of the spurious leaf expression observed in Figure 4.8-2B.

Alternatively, GUS expression in the torn leaf may signify stress-induced *KFB^T* expression. Staining was observed during flower abscission (Figure 4.8-2C) and this pathway is also tightly linked to invasive stresses such as wounding and pathogen attack (Taylor & Whitelaw, 2001). Endogenous GUS activity in plants has been observed with particular prevalence in floral reproductive organs (Alwen *et al.*, 1992). However, while Figure 4.8-2C shows carpel GUS staining, no activity was observed in the male reproductive organs of any flowers screened. Acidic pH and gibberellic acid treatment were both shown to induce intense GUS activity in most cell types from 23 model species, including *A. thaliana* (Sudan *et al.*, 2006). The process of flower abscission observed in Figure 4.8-2C is linked to raised levels of salicylic acid, jasmonic acid and gibberellic acid (Kim *et al.*, 2013b; Setyadjit *et al.*, 2006). Such acids associated with the abscission pathway could have provided the low pH required for exhibiting endogenous GUS activity in *Arabidopsis*, however this data is coherent with the detection of greatest *KFB^T* transcription levels from *P. vulgaris* gynoecea in earlier qPCR experiments (Figure 4.3).

These qPCR assays also detected *KFB^T* activity in *P. vulgaris* petals that decreased in the absence of *GLO^T* (Figure 4.3). Failure to observe *GUS* expression in *Arabidopsis* petals (that also lack *GLO^T*) further reinforces the hypothesis that *KFB^T* transcription in the second whorl necessitates activation by *GLO^T*. It also suggests that a non-*S* locus protein more common to higher plants may regulate *KFB^T* expression in the style, which successfully induced the GUS staining here (Figure 4.8-2C). Although post-seed dispersal expression of *KFB^T* would be too late for a role in *Primula* floral heteromorphy, this transgenic assay was carried out in a heterologous system and differences were expected.

Quantitative PCR assays were challenged by low overall levels of KFB^T expression and a genome-wide approach by Sun (*et al.*, 2007) also found that most Kelch genes were too lowly expressed to be reliably analysed. However, these temporal and spatial transcription analyses have allowed for speculation of several potential roles for KFB^T in floral heteromorphy. One interpretation of the data linked the presence of KFB^T with suppression of cell length. The temporal expression data (Figure 4.2) suggested the greatest transcription occurred in the first stage of flower bud development and the spatial analysis (Figure 4.4) showed KFB^T was particularly prevalent in the gynoecium at these early stages, especially stage two. Cell length of the gynoecium style structure is suppressed at this primary phase of flower bud maturation and this is the earliest observable facet of floral heteromorphy in thrum flowers (Webster & Gilmartin, 2006). Furthermore, KFB^T transcription was observed to significantly increase in the anthers at stages three and four of flower development (Figure 4.4). This is coherent with pollen maturation and release from the anthers. It follows that thrum pollen remains spherical but pin pollen grains present a longer oblong shape (Darwin, 1877), thus providing a second example of cell length suppression in thrum compared to pin that could be connected with KFB^T activity.

However, it can also be speculated that a positive correlation exists between KFB^T transcript levels and cell length. There is a reduction of KFB^T transcription in the petal whorl at stage three of thrum flower development (Figure 4.4) and it is here when cell length is suppressed in the upper corolla tube of thrums (Webster & Gilmartin, 2006). Short homostyle mutants displayed almost a total loss of KFB^T expression in the petals (Figure 4.3) and these flowers present anthers in the low position due to reduced length of anther filaments, which are enclosed within the petal whorl. Vastly increased transcription of KFB^T was measured in long homostyle gynoecia too (Figure 4.3), which presents a pin-height stigma due to the failed suppression of style length that usually occurs in thrum. Again, this suggests positive correlation between KFB^T activity and cell length. Moreover, while thrum pollen shape is suppressed, grain size is actually larger than in pin plants (Darwin, 1877). This could be related to the increase of KFB^T anther transcription in the latter half of flower maturation (Figure 4.4).

A third model implicates a role for KFB^T in the self-incompatibility system that reinforces floral heteromorphy. The greatest levels of KFB^T transcripts were mostly found in the gynoecium (Figures 4.3 & 4.4) and, although this is the site of style height dimorphism, it is also the site of pollen rejection in self-incompatibility. The quantity of KFB^T transcripts in

the gynoecium were only superseded by expression levels in the anthers at the final stage of development (Figure 4.4), which here release pollen; the other mediator of self-incompatibility. Expression of *KFB^T* in the reproductive organs was significantly greater in the short and long homostyles than in thrum (Figure 4.3). These two mutant lines are phenotypically similar only in their breakdown of the self-incompatibility system to become self-fertile.

Observations may therefore suggest a role in self-incompatibility and – though it initially appears counterintuitive that increasing *KFB^T* would lead to less stringency of the infertility system – this could be explained in two possible ways. The first is that *P. vulgaris* may reject all pollen by default and *KFB^T* facilitates the breakdown of an unknown molecule to enable compatibility. This would make infertility the default state and receptivity to pollen the exception. Secondly, RNA-Seq data (Figure 3.6.3-3) suggested the majority of *KFB^T* transcripts in the homostyle flowers may undergo cleavage via post-transcriptional modification and therefore be non-functional. Raised *KFB^T* expression may therefore not signify increased *KFB^T* functionality. Failed translation of cleaved *KFB^T* transcripts could inhibit a currently unidentified negative feedback system that usually represses *KFB^T* expression. In the event of this malfunctioning feedback mechanism, dysfunctionality of *KFB^T* would lead to increased transcript levels, as observed in the homostyles (Figure 4.3). On the contrary, no *KFB^T* transcripts were detected in styles of *P. oreodoxa* (Zhao *et al.*, 2019), which is a strongly self-compatible species (Yuan *et al.*, 2019). Transcripts of *CCM^T* and *PUM^T* were also absent (Zhao *et al.*, 2019).

The short homostyle phenotype has previously been connected to a mutation in the *GLO^T* transcription factor from the *S* locus (Li *et al.*, 2016). This corresponds with a reduction of *KFB^T* expression in short homostyle petals (Figure 4.3), which suggests *GLO^T* may indirectly cause the short homostyle phenotype by acting through *KFB^T*; the knockdown of which could lead to suppressed anther filament length via targeted degradation in a downstream pathway. The substantial increase of *KFB^T* in the gynoecia of long homostyle flowers also suggested a link between *KFB^T* and *CYP^T*, mutation of which is believed to be responsible for the long homostyle phenotype (Huu *et al.*, 2016). Although this work revealed a reduction of brassinosteroid content following the loss of *CYP^T* function, no direct mechanism for this relationship was proposed. Alternatively, it has been proven that the BSU1 Kelch-repeat protein modulates brassinosteroid response in *Arabidopsis* (Mora-

García *et al.*, 2004). It could therefore be KFB^T that directly leads to the long homostyle phenotype via its relationship with CYP^T .

It is known that some Kelch proteins can bind to several targets and therefore exhibit multiple functions (Kim *et al.*, 2013a; Zhang *et al.*, 2013). It was thus proposed in chapter three that the timing and location of KFB^T production may be tightly regulated to limit exposure of target proteins and restrict spurious degradation activity. The qPCR assays have provided a map regarding the temporal and spatial transcription patterns of KFB^T . These assays have presented areas of focus for downstream experiments. However, it is unlikely that KFB^T could simultaneously carry out all three functions proposed here due to their contradictory nature. For example, interpretations of the data correlate KFB^T both positively and negatively with cell length. A yeast-two-hybrid experiment was therefore designed to identify interacting partners of KFB^T and highlight potential protein targets to infer which pathway it modulates (Chapter 6). Vector constructs were also assembled to generate transgenic loss-of-function and gain-of-function mutant lines to elucidate the role of KFB^T in floral heteromorphy (Chapter 5).

Chapter 5

5 The Function of *KFB^T*

Bioinformatics had been used to analyse the *KFB^T* sequence (Chapter 3) alongside temporal and spatial expression analyses at the transcript and protein levels (Chapter 4). Subsequent *in vivo* misexpression experiments were utilised to investigate the function of *KFB^T*. Two knockdown and two overexpression methods were used to modify *KFB^T* function and provide data regarding its role in floral heteromorphy.

5.1 Comparison of *KFB^T* in Self-Compatible and Self-Incompatible Plants

The aspects of floral heteromorphy are easily observed by eye or with a light microscope. Transgenic *KFB^T* mutants were therefore expected to present distinguishable phenotypes that help to elucidate gene function. However, *Primula* heterostyly is reinforced by a self-incompatibility system that makes a plant unreceptive to its own pollen (Darwin, 1877). Expression of *KFB^T* had been detected in floral reproductive organs (Figures 4.3, 4.4 & 4.8-2), which are the dimorphic sites of heterostyly but also the mediators of self-incompatibility. If *KFB^T* functions in this self-incompatibility system, mutants may fail to present a visually obvious phenotype. A preliminary PCR screen was therefore used to test this hypothesis prior to carrying out the downstream transformation assays.

Self-compatible mutant lines of *P. veris* had been identified from a glasshouse population that were part of a forward genetics experiment in which seeds were exposed to fast neutron bombardment at the International Atomic Energy Agency in Vienna, Austria (Li, unpublished). The radiation caused random mutations throughout the genomes of the seeds, which were grown and self-crossed to screen for modifications of the self-incompatibility phenotype.

PCR (Chapter 2.9) confirmed there were no size or sequence differences between the *KFB^T* amplicons from self-compatible and self-incompatible *P. veris* plants. However, this could not conclusively eliminate *KFB^T* as a candidate gene in the self-incompatibility system. An assay from Dr Jinhong Li found *KFB^T* transcripts to be almost undetectable in *P. veris* cDNA samples (Li, unpublished). This gene may have a vital role in self-infertility but another gene, either upstream or downstream in the pathway, could instead carry the mutation

that resulted in loss of self-incompatibility. This cannot be confirmed until the network that mediates self-incompatibility has been elucidated.

5.2 A KFB^T Constitutive Overexpression Vector

Overexpressing a gene of interest is a method commonly used in functional analyses to observe the mutant phenotypes displayed by exaggerating the effect of a gene in a target host (Prelich, 2012). This may provide strong evidence for the role of that gene within the organism. These assays may cause visible physiological changes to the host that are immediately obvious but sometimes the products of pathways and cascades must be measured to observe more subtle modulations in gene networks.

An overexpression vector was therefore made with pBRACT114 (Chapter 2.11.3) using the KFB^T open reading frame under regulation of a 35S promoter derived from the Cauliflower Mosaic Virus (Figure 5.2). This promoter causes constitutive overexpression of the gene of interest throughout the entire transgenic plant.

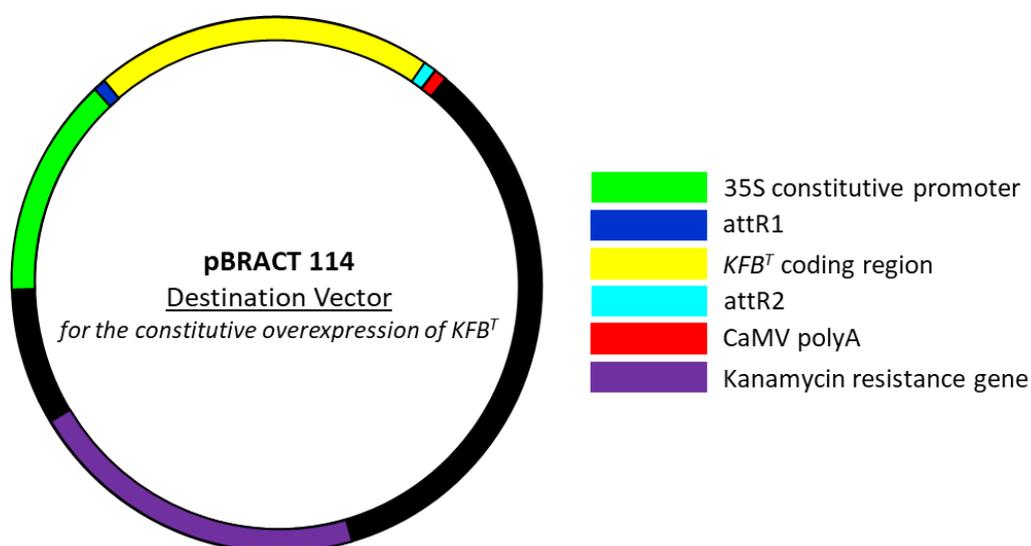


Figure 5.2: A pBRACT114 construct was assembled with the 35S promoter for constitutive overexpression of the KFB^T coding region.

5.3 A Vector Containing *KFB^T* and Its Native Promoter

In addition to using the 35S Cauliflower Mosaic Virus promoter for constitutive overexpression, a vector containing *KFB^T* under regulation of its native promoter was also assembled (Figure 5.3). This allows *KFB^T* expression at native levels instead of those from a constitutive promoter, which may have a deleterious effect on the host in such high quantities. Use of a native *P. vulgaris* promoter also removes the potential risk of the 35S promoter from a Cauliflower Mosaic Virus being ineffective in *Primula*.

The construct was designed for incorporation into the genomes of plants that do not naturally possess *KFB^T*, such as *Arabidopsis* or pin-form *Primula* species. This plasmid contains a kanamycin resistance gene and has the components required for replication in bacteria. This vector was transformed into *Arabidopsis* due to the rapid life cycle of this species and its well-optimised techniques (Koornneef & Meinke, 2010; Somssich, 2019). This tested if *KFB^T* could act in other higher plants or whether it requires a more *Primula*-specific gene for transcriptional activation or downstream function.

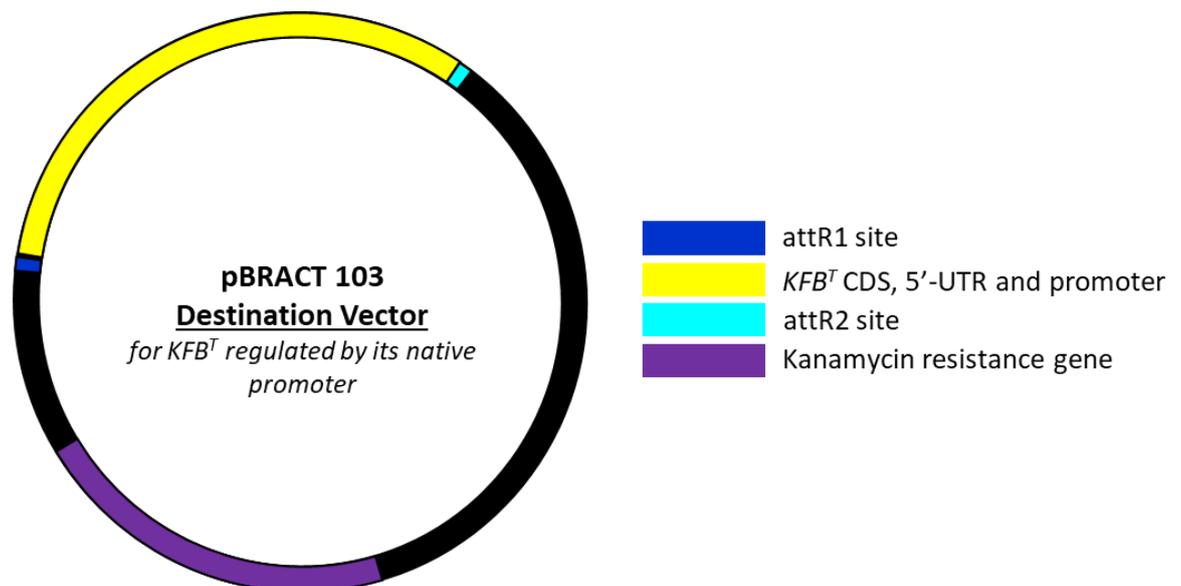


Figure 5.3: The pBRAC103 vector was used to generate a construct containing *KFB^T* under regulation of its native promoter and 5'-UTR regions.

5.4 A KFB^T RNAi Knockdown Vector

To complement data from the overexpression studies, an RNAi knockdown vector was assembled (Chapter 2.11.4) to investigate the phenotypic effects caused by reducing KFB^T transcript levels in transgenic plants. These mutagenesis experiments provide functional data to aid elucidation of the KFB^T role in floral heteromorphy. Although a viral method (Chapter 5.7) can present data within two weeks (Baulcombe, 1999), the target gene is only knocked down in successfully infected tissues. It does not eliminate the effects of target gene activity prior to infection and the genome itself remains unmodified.

To obtain stable transgenic lines that inherit the mutation through generations, an RNAi construct was thus made for KFB^T . This vector contains two inverted duplicate fragments from the 3'-end of the KFB^T transcript that are joined by a linker in opposing orientations (Girin *et al.*, 2010), thereby folding into a double-stranded hairpin loop that triggers the RNA interference pathway and leads to the degradation of KFB^T transcripts.

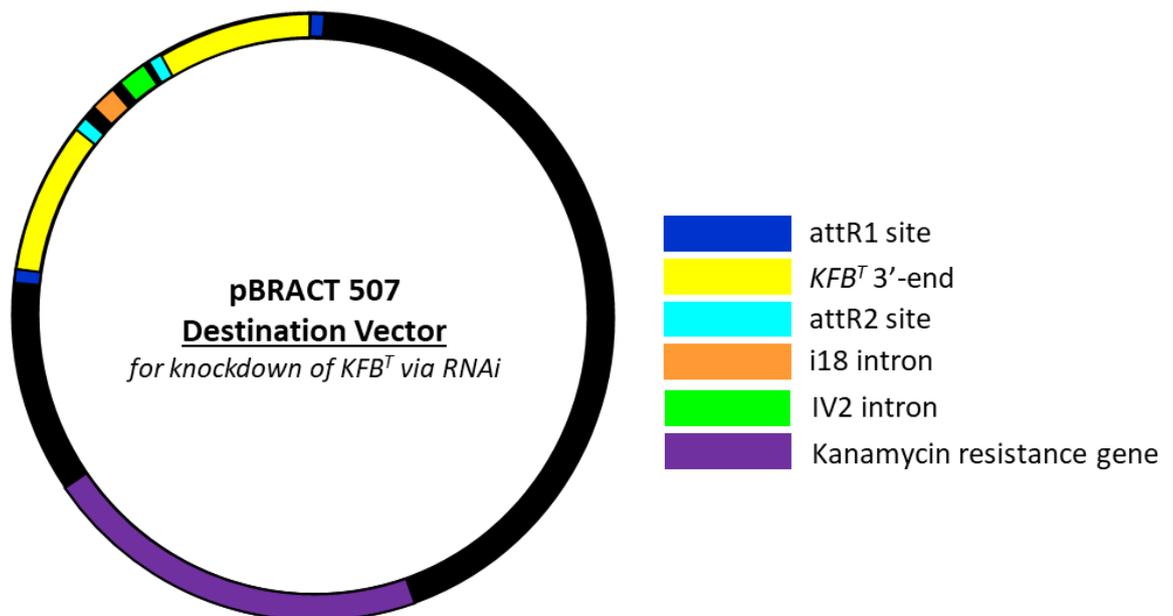


Figure 5.4: The construct assembled for the knockdown of KFB^T via RNA interference.

5.5 Screening for Transgenic *Primula*

The series of KFB^T vectors were transformed into *P. vulgaris* (Chapter 2.15.1), which subsequently underwent selection and tissue culture (Chapter 2.15.2). Transformed leaf sections were monitored for survival on kanamycin and growth of callus tissue. Successful transformants would possess an active kanamycin resistance gene on the destination

vector. Cells without this plasmid would turn colourless. Most leaves turned transparent within a week, which was indicative of unsuccessful transformation of those samples. Other leaf tissue hardened – displaying early signs of callus formation – but browned slowly over time. Some leaves appeared to retain their colour for up to 6 months but did not form callus. Numerous repeats of this experiment had been carried out across two years. Overall, all tissue died, either due to unsuccessful incorporation of the kanamycin resistance gene into the *Primula* genome or because of failed transgenic tissue culture.

In mid-2019, Dr Sadiye Hayta localised the source of this problem to the use of Phytigel™ (unpublished). Two years previously, Sigma-Aldrich had changed the recipe of this gelling agent and the tissue culture method had since been ineffective, both in *Primula* and in Barley (unpublished). Agarose was trialled instead but this proved inadequate. A gelling agent called Gellan Gum from Alfa Aesar™ is the currently proposed solution to the problem. There was insufficient time remaining in the PhD project to repeat these experiments and so it was decided to fully divert focus toward *Arabidopsis* transformation instead. However, this series of *KFB^T* destination vectors is available to colleagues and will be essential in carrying out the future transgenic experiments that are mandatory for understanding the mechanisms underpinning floral heteromorphy.

5.6 Screening for Transgenic *Arabidopsis*

The *KFB^T* plasmids containing the overexpression and native promoters were also transformed into *Arabidopsis* (Chapter 2.16). This allowed faster results and circumnavigation of the issues encountered in transforming *Primula*. A total of eleven *Arabidopsis* lines overexpressing *KFB^T* via the 35S constitutive promoter (Figure 5.2) were investigated. There were four flowers from three plants that had grown an extra petal and sepal (Figure 5.6A). Eighteen transgenic lines containing *KFB^T* under regulation of its native promoter (Figure 5.4) were also investigated, with a further eight third-generation plants from one additional line. One of these plants presented three flowers containing only five anthers (Figure 5.6C) and another four flowers from three plants had only four stamens (Figure 5.6D); one of which had a fifth with retarded development (Figure 5.6B). This was also observed once in the transgenic *GUS* plants (Chapter 4.7; not shown), which contain only the *KFB^T* promoter region and should not demonstrate any modified function, thereby suggesting this was a developmental anomaly and not related to *KFB^T* activity.

Specimens with the additional sepal and petal were found only in the constitutive overexpression line, however the low number of mutant flowers suggests this is also unrelated to *KFB^T* activity. As every cell should carry the same transgenic genotype, an authentic mutant would be expected to occur with more consistency across flowers.

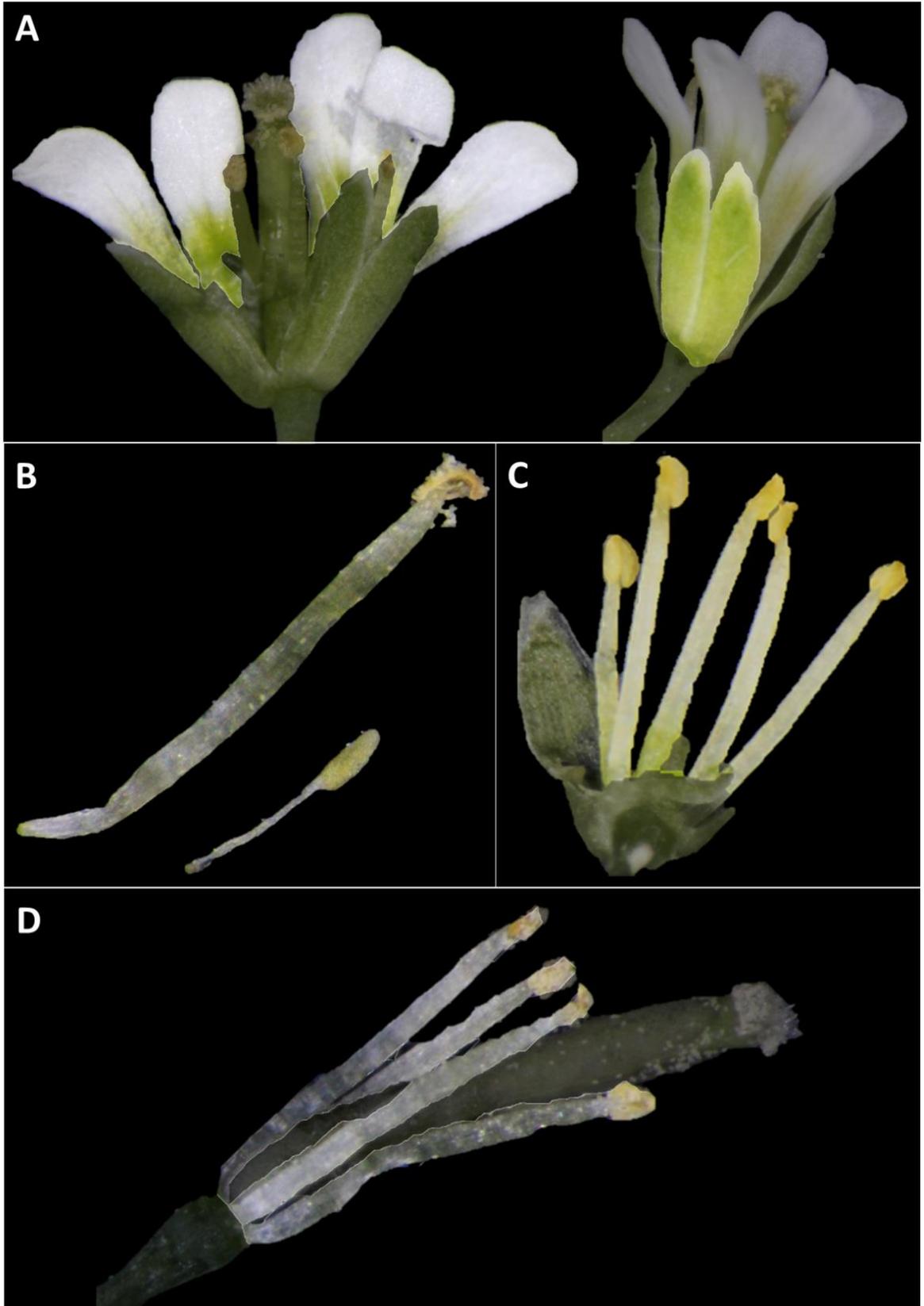


Figure 5.6: Transgenic *Arabidopsis* flower tissue transformed with *KFBT* overexpression vectors. A) A mutant flower displaying an extra sepal and petal. B) Two transgenic stamens; one presenting retarded development. C) A mutant flower containing only five stamens. D) A mutant flower containing only four stamens. Flowers were dissected for clarity and image brightness has been edited to highlight appropriate flower parts.

5.7 Virus Induced Gene Silencing of *KFB^T*

Virus Induced Gene Silencing (VIGS) was used to knockdown *KFB^T* in *Primula*. The RNAi vector (Chapter 2.11.4) required *Agrobacterium*-mediated transformation into the *Primula* genome, with subsequent callus formation through tissue culture, further growth and selection of second-generation seeds. Alternatively, VIGS vectors are not incorporated into the host genome and only require infection of the virus in order to knockdown transcripts, thus yielding results far more rapidly (Baulcombe, 1999).

This vector was electroporated into *Agrobacterium* (Chapter 2.14.2), of which cells were injected into the stems of Primroses (Chapter 2.17). The virus then generates short double-stranded transcripts (Unver & Budak, 2009) from the 3'-end of *KFB^T* that trigger an immune response in the plant (Voinnet, 2001) that consequently knock down all *KFB^T* transcripts present. In this way, *KFB^T* is disabled and the effect of minimising expression of this gene can be observed and its function deduced.

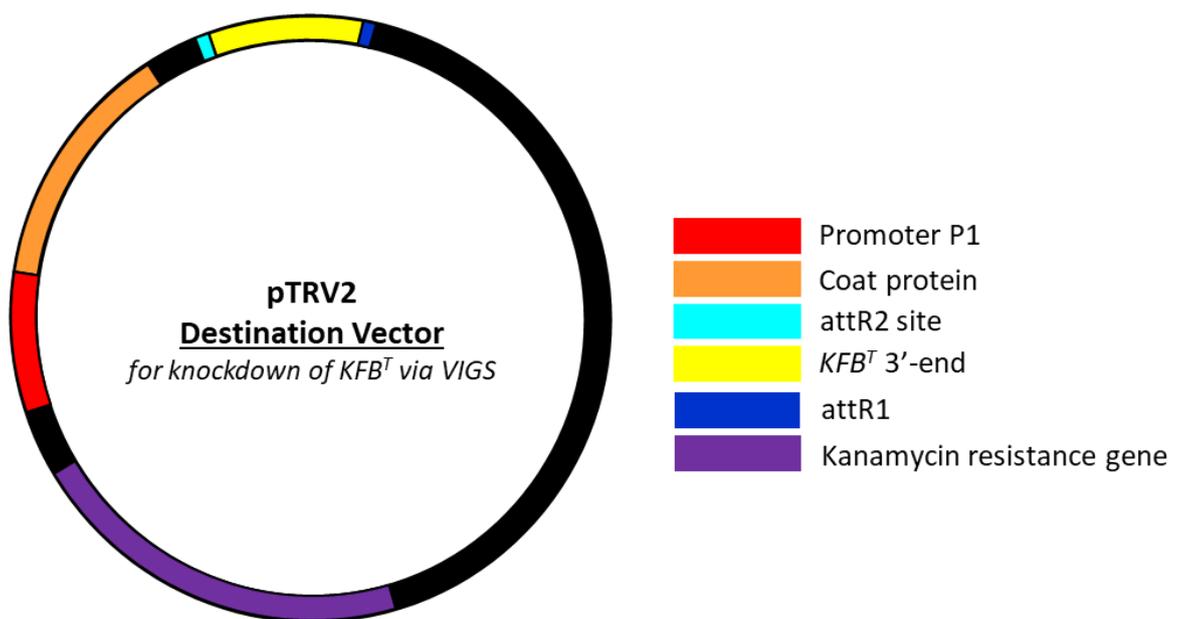


Figure 5.7-1: The pTRV2 vector was used in combination with pTRV1 for the virus-induced gene silencing of *KFB^T*.

Approximately 10 % of treated thrum plants presented at least one flower with an abnormal corolla tube (Figure 5.7-2). These altered corollas were curved and the most extreme samples were warped into an S-shaped crinkle. Pin and thrum samples that underwent the negative control treatment did not present any irregularities.

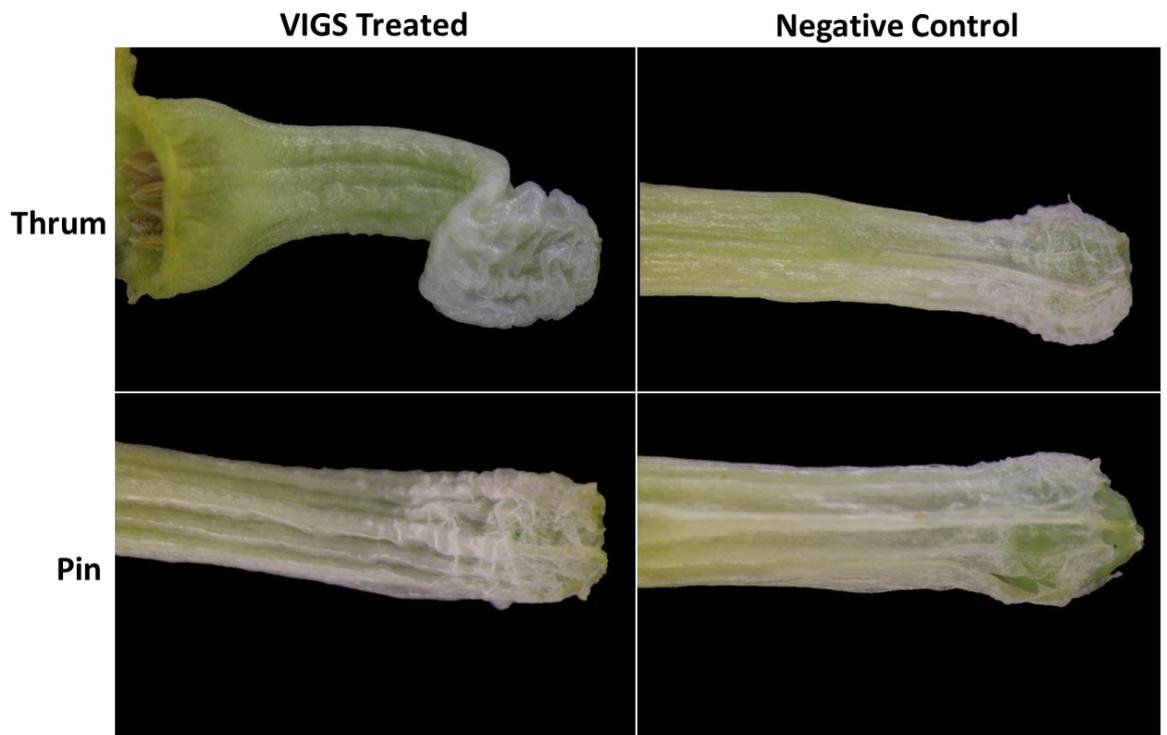


Figure 5.7-2: Abnormal phenotype in thrum corollas following viral induced gene silencing of *KFBT*. The S-shaped crinkle was not observed in treated pins or plants that underwent injection with a negative control treatment.

These abnormal plants could also be identified from the face of their flowers. The corolla mouth is flush with the petal surface in wildtype Primroses (Figure 5.7-3, right). However, in abnormal samples, the petal surface had retracted from the mouth of the flower due to the lower flower tube crinkling and resulted in protrusion of the corolla tip housing the anthers (Figure 5.7-3, left). Forty thrum plants underwent treatment in this first VIGS assay and five were used as negative controls. Following observation of the low rate at which plants presented altered phenotypes (~10 %), it was decided that a second attempt would be carried out utilising near-equal numbers of VIGS-treated and negative control specimens.



Figure 5.7-3: Abnormal flower mouth following viral induced gene silencing of KFB^1 . Left) Treated thrums presented flowers with protruding corolla openings and retracted petal surfaces linked to warping of the corolla tube beneath. Right) The flower mouth of wildtype Primroses is flush with the petal surface.

A total of twenty thrum plants were used for VIGS analysis with the *KFB^T* construct and fifteen were used as negative controls. Additional flowers with warped corollas were obtained from this second VIGS assay. A novel abnormality was also observed in the fusion of the sepal to the corolla along the inner surface of the curved flower tubes (Figure 5.7-4).

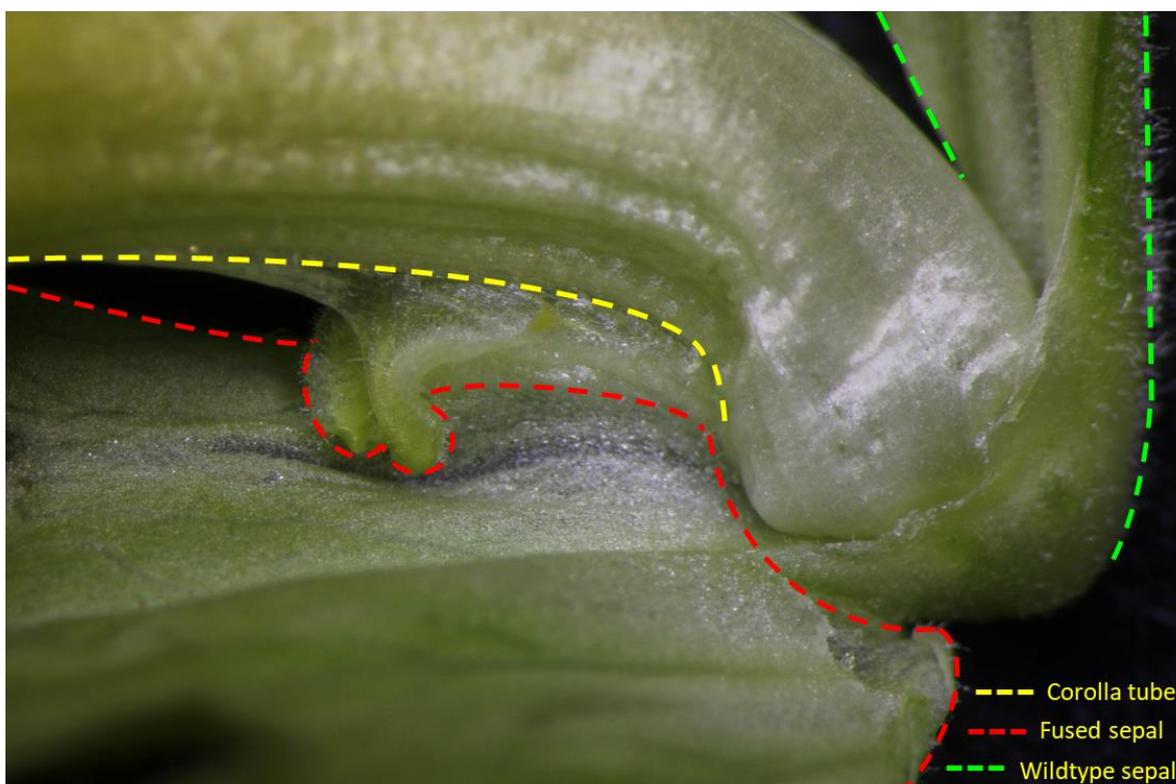


Figure 5.7-4: Abnormal corolla tube base from a *Primula* plant treated with a construct to virally induce gene silencing of *KFB^T*. The flower tube shows webbed tissue fusing the sepal (red line) and petal (yellow line) whorls along the inside edge of a curved corolla. A non-fused wildtype sepal is outlined behind (green line).

A potential relationship between *KFB^T* and the *GLO^T* transcription factor from the *Primula* *S* locus had been indicated by observing loss of *KFB^T* transcription in petals of short homostyles (Figure 4.3) that corresponded with loss of *GLO^T* function (Li *et al.*, 2016). To investigate this further, a construct assembled by Dr Sadiye Hayta to silence *GLO^T* was introduced into five thrum plants (2016, unpublished). These plants also presented the warped corolla with fusion between the sepal and petal whorls (Figure 5.7-5). Furthermore, the edge of the petal blade had adopted green sepal identity following the seam of the petal-sepal fusion (Figure 5.7-5C), indicating loss of B-function floral identity (Coen & Meyerowitz, 1991).



Figure 5.7-5: Irregular flower tissue following viral induce gene silencing of *GLO7*. A) Fusion between the sepal and petal blade was observed. B) Abnormally curved corollas were fused to the sepal; the webbed tissue has here been manually split. C) The edge of the petal blade adopted green sepal identity following the seam of petal-sepal fusion.

At this stage, it was noted that some of the negative control flowers also exhibited severely bent corollas (Figure 5.7-6). The Polyanthus had regularly presented twin flowers emerging from a single calyx, which was viewed as a horticultural anomaly. The corollas of these twin flowers curved sharply away from each other. Although the VIGS samples shown were not double flowers, it was unclear whether this curving was a milder form of the twin flower phenotype or if it was truly caused by *KFB^T* silencing. The sepal-petal fusion was not observed in negative control samples. It was therefore uncertain whether this fusion could be a separate phenotype caused by knockdown of *KFB^T* or another facet of the complex phenotype consisting of protruding flower mouths and warped corollas now believed to be a product of horticultural breeding.



Figure 5.7-6: Negative control flowers from an experiment to virally induce gene silencing of *KFB^T* presented warped corollas (A &B) and recurved flower mouths (C). The background was removed for clarity and some sepals were excised to reveal the crinkled corolla.

A third repeat of the VIGS experiment was executed to gain further evidence of the abnormal phenotypes and confirm that petal-sepal fusion was linked to *KFB^T* silencing while the other phenotypes were caused by variation in the horticultural breed. This final assay also served to assess whether irregularities in the previous negative controls could have been caused by accidental cross-contamination of the *KFB^T* viral vector. However, the disfigured corollas were again observed in negative control samples; more so than in the treated population.

The negative controls also displayed fusion between the sepal and petal whorls. Therefore, *KFB^T* activity could not be associated with these phenotypes. Self-crosses were carried out to screen for breakdown of the self-incompatibility system but the Polyanthus appeared to be naturally self-fertile anyway. Pollen grains and styles were also compared between treated and non-treated samples (not shown) but no morphological differences were identified.

5.8 Discussion

To help understand gene function, vectors to induce misexpression of *KFB^T* were constructed and transformed into plants. An RNAi knockdown vector was assembled but stable transformation into *P. vulgaris* failed. Successful transformation was achieved in *Arabidopsis* with *KFB^T* under regulation of the 35S constitutive promoter in addition to a second vector containing its native promoter and 5'-UTR regions. The virus induced gene silencing of *KFB^T* was carried out in Polyanthus plants as a second method of functional analysis via gene knockdown.

Alterations in floral organ number were observed when *KFB^T* was expressed in *Arabidopsis* (Figure 5.6), with the gain of a petal and sepal in some plants and the loss of up to two stamens in others. These organ numbers have been observed when alleles of the *PERIANTHA* (*PAN*) transcription factor were mutated; 84 % of *pan-1* mutants presented five sepals by day 45, 70 % had five stamens and 72 % of the first fifteen flowers had five petals (Running & Meyerowitz, 1996). Mutations in another *Arabidopsis* gene encoding a Kelch F-box protein, *UNUSUAL FLORAL ORGANS* (*UFO*), led to modified stamen and petal number via action with *LEAFY* to co-regulate *APETALA3* and *PISTILLATA* (Levin & Meyerowitz, 1995).

However, the low number of mutant flowers from this KFB^T screen (just eleven flowers from almost thirty transgenic plants) was consistent with natural variation of floral organ number in wildtype *Arabidopsis* plants. Although a typical *Arabidopsis* flower has six stamens, the mean number is 5.93 due to the loss of one or two in some wildtype inflorescences (Penin & Logacheva, 2011). Acquisition of an additional petal and sepal is also known to occur less frequently – typically in basal flowers – and is caused by increased floral meristem size induced by the transition from vegetal to floral growth (Penin & Logacheva, 2011). These correspond to the organ numbers observed in this KFB^T screen (Figure 5.6).

Environmental variables are also known to cause these floral organ number modifications, such as being moved from non-flowering to flower-inducing conditions (Penin & Logacheva, 2011). Lower temperatures correlate with increased petals and fewer stamens in *Rosa chinensis* (Han *et al.*, 2018). The reduced stamen from Figure 5.6B was also noted in *Arabidopsis* plants from the *GUS* reporter assay without transgenic KFB^T functionality (Chapter 4.8; not shown) and these lines behaved as a negative control to further confirm modified KFB^T activity was not the source of this variation.

Use of the 35S constitutive promoter and the previously observed *GUS* expression in transformed *Arabidopsis* gynoecia (Figure 4.8-2) confirms that the lack of a significant mutant phenotype is not due to a lack of KFB^T transcription. Therefore, it may instead be suggested that KFB^T requires a downstream target protein that is not present in *Arabidopsis* – perhaps even a partner protein from the *Primula S* locus. Alternatively, transgenic expression may simply commence too late in development for a mutant phenotype to be observed (Chapter 4.8). When *Primula* transformation issues are resolved and a stable protocol has been established, insertion of KFB^T into the pin genome will confirm whether it is a general housekeeping protein or an *S* locus member that KFB^T requires. An alternative possibility is that KFB^T permits pollen-stigma compatibility between *Primula* plants but no difference was observed in *Arabidopsis* because it is already self-compatible.

A VIGS assay was carried out on *Primula* (Chapter 5.7) to knockdown KFB^T functionality in thrum plants. It initially appeared that anther filaments had shortened in these flowers. As these filaments are fused between the endodermal and epidermal cell layers of the petal whorl, their shortening would pull down the corolla and result in the warping that was observed here (Figure 5.7-2) – if length of the flower tube itself remained unaffected. This is equivalent to the mechanism believed to cause short homostyle phenotypes, which

present a concertina-like crinkling of the corolla when viewed under a microscope (Li, unpublished). Abnormal phenotypes appeared in ~10 % of treated plants and, likewise, only 2 of 22 *Primula* that underwent VIGS treatment in the Huu (*et al.*, 2016) experiment resulted in abnormalities to reveal a function in style length suppression for *CYP^T*.

Plants that underwent VIGS treatment to silence *GLO^T* activity also demonstrated crinkled corolla phenotypes (Li, unpublished). Short homostyles have previously been linked to *GLO^T* mutations (Li *et al.*, 2016). This appeared to support the potential link between *GLO^T* and *KFB^T* that was proposed by qPCR analyses, in which *KFB^T* transcription was lost in the petal whorl (Figure 4.3). This observation suggested that *KFB^T* might be directly involved in mediating heteromorphic anther position in *Primula*. One possible conclusion could be that *GLO^T* is the transcription factor that activates *KFB^T* expression, which would explain there is no *KFB^T* expression in short homostyle petals that have lost *GLO^T* function. This in turn could mean that loss of *GLO^T* indirectly regulates anther height through *KFB^T*.

A second phenotype linked to the crinkled corollas from VIGS assays was observed. While the flower mouth is flush with the petal surface in wildtype plants, the abnormal VIGS flowers presented a corolla opening that protruded from withdrawn petals (Figure 5.7-3). This was likely a secondary effect of the retracting corolla beneath, though this has not been observed in short homostyles.

Viral silencing exhibits knockdown of genes proportional to dosage of the pTRV vectors; it is not a true knockout. Tissues with an increased quantity of successfully infected cells would thus be expected to demonstrate greater effects. With respect to the abnormally curved corollas, it was therefore predicted that the anther filament along the inside axis of the curve must have been exposed to greater *KFB^T* silencing by the VIGS vectors and therefore pulled on the corolla with greater tension, thereby causing the flower to be curved in that direction.

Alternatively, a second explanation for this curvature became evident in the second VIGS assay (Figure 5.7-4). Growth of the corolla was restricted by webbed tissue that fused the sepal and petal whorls along one side. In plants treated with VIGS constructs to silence *GLO^T*, it was further noted that the petal blade edge had adopted green sepal identity following the seam of this fusion (Figure 5.7-5). However, these malformations also appeared in some negative control samples too (Figure 5.7-6).

It was therefore impossible to conclusively link *KFB^T* function with these irregular phenotypes, which may instead be a result of the extensive inbreeding and hybridisation that produce such commercial lines. The Polyanthus horticultural variety was selected for its thick and long flower stalk that is not present in wildtype *P. vulgaris* but is well-suited to the VIGS injection procedure. The *P. veris* species presents an umbel of flowers atop a significantly longer flower stalk than *P. vulgaris* but is most likely too thin for adequate VIGS injection. Breakdown of the self-incompatibility system was another difference observed in the Polyanthus variety (Chapter 5.7). Huu (*et al.*, 2016) successfully utilised *P. forbesii* in a VIGS experiment to investigate style length.

Though the five *S* locus genes have been identified, much work remains to elucidate the network and pathways that underpin floral heteromorphy. Although several questions can be answered bioinformatically, such as the mechanism responsible for maintaining integrity of the hemizygous *S* locus and uncovering the chronological sequence of events that led to the evolution of heteromorphy, an extensive focus on transgenic approaches is required. It has been found that the silencing effect of VIGS treatments can persist for over two years and may even be transmitted to seedlings (Senthil-Kumar & Mysore, 2011). In the current absence of a reliable *Primula* transformation system, VIGS may offer a feasible workaround. This would allow for the treatment of seedlings in advance of their transition to reproductive growth, which would increase the quantity of infected cells and mean the plant would endure targeted gene silencing across the entirety of flower development.

Even with a routine transformation protocol, the slow life cycle and annual flowering period of *Primula* would remain a hurdle. Time could be invested to establish the infrastructure and methods that would provide a rapid turnover or continuous supply of *P. vulgaris* flowers, but it would be difficult to glean data from plants subjected to hormonal treatments or modified lighting and growth conditions. While the necessity for transgenic *Primula* cannot be replaced, work can be assisted via the transformation of *Arabidopsis*.

Chapter 6

The Operation of KFB^T

In addition to deducing the function of KFB^T , experiments were also carried out to explore how KFB^T executes this function. A yeast two-hybrid assay was used to screen *Arabidopsis thaliana* and *Primula vulgaris* cDNA expression libraries to identify partner proteins of KFB^T . As KFB^T has predicted involvement in targeted protein degradation, using yeast two-hybrid to elucidate the protein it binds to would identify which protein is targeted for degradation. Further research on target proteins could highlight the pathway modulated by KFB^T and offer more information about its operation and function.

Yeast two-hybrid assays work by using a *Saccharomyces cerevisiae* yeast cell carrying two vectors; a prey construct containing the coding sequence for a transcription activation domain fused to cDNA from the tissue to be screened and a second bait construct with the coding sequence for a DNA binding domain fused to the cDNA for the protein of interest (Chien *et al.*, 1991). These two protein domains must be expressed in the cell and if they can be brought together through a protein-protein interaction between the bait and second prey protein, they can activate expression of a histidine biosynthesis gene (or other reporter gene) via the upstream activating sequence (Fields & Song, 1989). The cell cannot survive without this essential amino acid and so only those containing successfully interacting protein partners grow to form colonies.

This approach was used to screen cDNA hybrid expression libraries from *Primula* and *Arabidopsis* (Matchmaker™, 2007). Transcripts from a cDNA pool were ligated into a prey vector with the GAL4 transcription activation domain and transformed into yeast strain AH103 (Chapter 2.20.2). The bait construct (Chapter 2.11.2) containing KFB^T ligated in frame to the sequence encoding a GAL4 DNA binding domain was transformed into yeast strain Y187 (Chapter 2.13). These two compatible yeast strains were mated to result in diploid cells containing both plasmids (Chapters 2.20.2 & 2.20.3). Mated cells were grown on media lacking histidine and only colonies containing plasmids encoding fusion proteins that can interact may survive. The prey plasmid from surviving colonies can then be isolated and sequenced to identify the KFB^T binding partner. A *P. vulgaris* flower cDNA library was the primary focus of this experiment but an *A. thaliana* cDNA library was also tested.

6.1 Preparation of *KFB^T* Activation Domain and Binding Domain Vectors

Although only the *KFB^T* bait construct was used to screen the prey vector protein library, the *KFB^T*-AD vector was also made for future testing of interactions in both directions. These vectors positioned the fusion domain at the 5'-end of the inserted *KFB^T* fragment. This is the same terminus as the *KFB^T* F-box domain. An alternative plasmid, such as pGBKCg, would have fused the binding domain to the 3'-end of *KFB^T* instead. However, it was decided to use pGBKT7 because the fused domain could potentially block binding activity of the 5'-end F-box. This was favourable because finding binding targets of the Kelch domain was the true experimental aim. The already well-characterised F-box functions in recruitment to the SCF complex, therefore identifying its binding partners was irrelevant and would only have interfered with this assay.

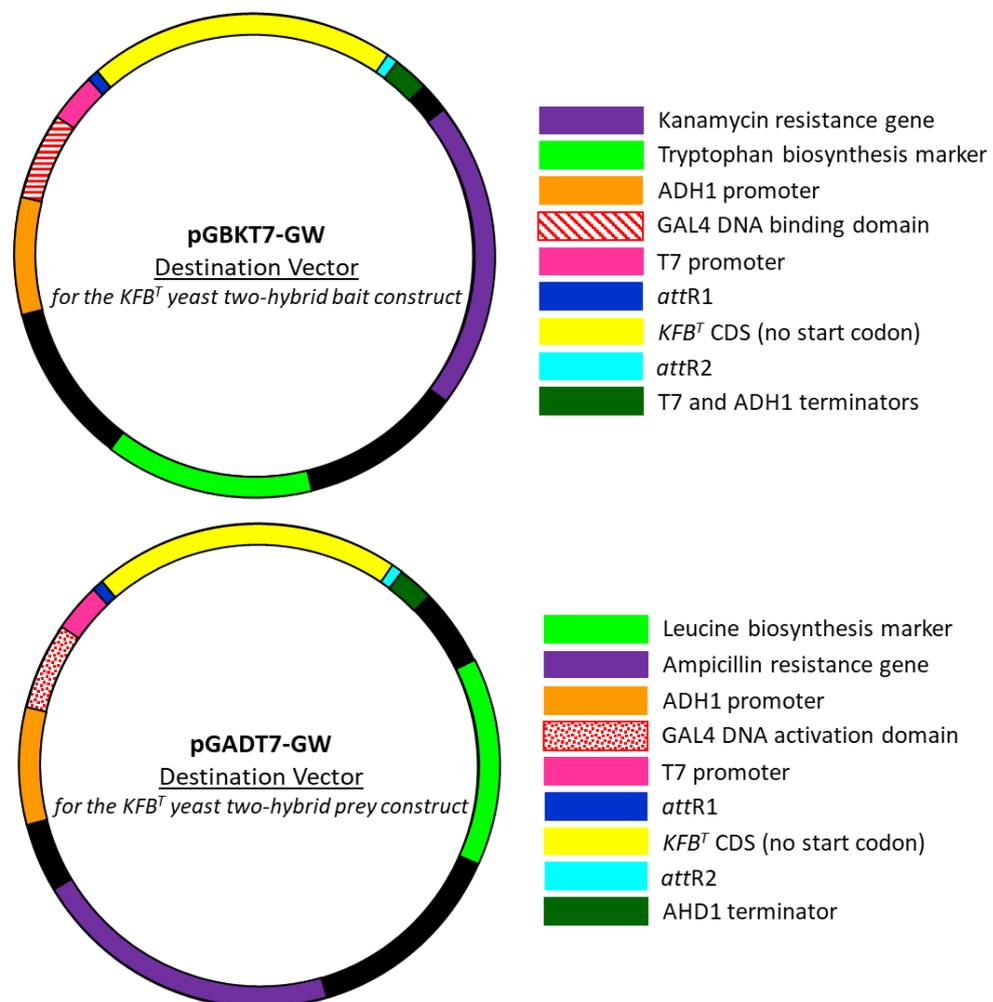


Figure 6.1: Vector maps for *KFB^T* yeast two-hybrid plasmids. Bait and prey constructs were both made with the *KFB^T* coding sequence for downstream cDNA library screens and protein-protein interaction tests. The pGBKT7 bait plasmid contains the tryptophan biosynthesis marker for selection and the GAL4 DNA binding domain. The pGADT7 plasmid contains the leucine biosynthesis marker and the GAL4 DNA activation domain. These two domains are brought together by protein-protein interaction to activate expression of the histidine reporter gene for survival.

6.2 Screening for Specific Interaction Between GLO^T and KFB^T

The qPCR assays had pointed toward a potential link between loss of GLO^T expression and loss of KFB^T expression (Chapter 4.3). This was presumed to take place at the transcript level – because GLO^T is a transcription factor (Kent, 2016) that may regulate activation of KFB^T expression – but this pair were also tested for interaction at the protein level, which would potentially signify a negative feedback loop. The PvGLO and PvDEF proteins were also included as positive controls because they are a known interacting pair.

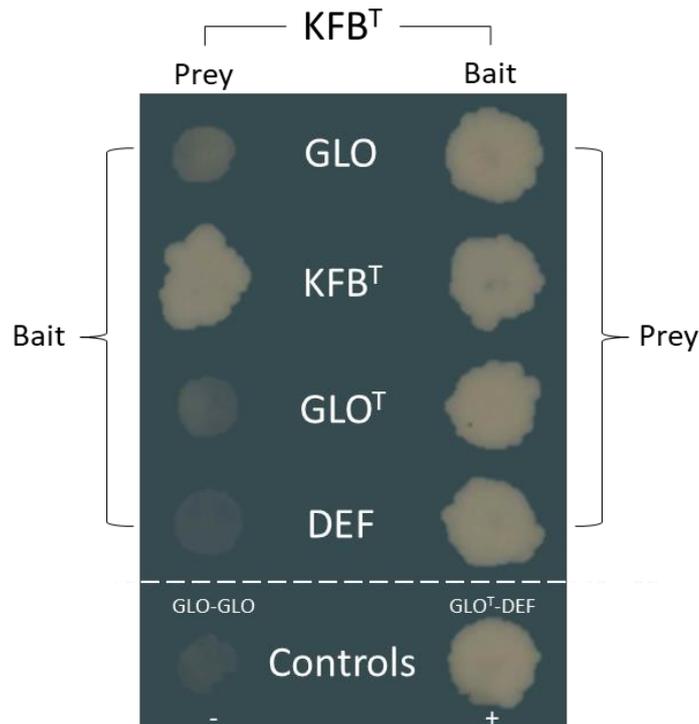


Figure 6.2-1: Protein-protein interaction screen between KFB^T with GLO , GLO^T and DEF . Growth indicates successful protein binding and activation of the histidine reporter gene required for survival. Homodimerisation of KFB^T was observed and KFB^T from the prey construct bound with all tested proteins. Known binding partners, PvDEF and PvGLO, were used as a positive control. Homodimerisation of PvGLO is known not to occur and so was here used a negative control.

Cells containing the KFB^T bait construct and the GLO^T , $PvDEF$, KFB^T or $PvGLO$ prey constructs all grew on -WHL dropout media (Figure 6.2-1). This shows that these proteins interacted successfully with KFB^T to activate expression of the histidine biosynthesis gene for survival. However, interactions in the reverse direction between the bait samples against the KFB^T prey did not occur. The positive control interaction between PvGLO and PvDEF functioned correctly. The PvGLO protein does not homodimerize (Causier, unpublished) and the PvGLO-PvGLO negative control mating correctly failed to grow on the -WHL dropout media. All results were consistent at 10 mM of 3-AT, thereby demonstrating stable affinity at the highest level of inhibitor tested.

Identifying such unidirectional interactions is common in yeast two-hybrid assays and they are usually denoted to steric hindrance (Uetz & Dong *et al.*, 2006). It is unclear why the interaction did not work bidirectionally in this case. The problem cannot simply be linked to interference of the fused activation domain from the *KFB^T* prey construct because the *KFB^T* self-mating assay demonstrated successful dimerization. This would suggest that interference instead came from the binding domains on the *PvGLO*, *GLO^T* and *DEF* constructs. However, these have elsewhere functioned correctly in unpublished experiments by Dr Causier. The complex conformations that arise when *KFB^T* binds to these proteins must therefore hinder the activation or binding domains.

This experiment highlighted potential interaction between *KFB^T* and *GLO^T*, *GLO*, *DEF* and itself. The *GLO^T* gene is expressed in the fused petal and anther whorls of *P. vulgaris* (Kent, 2016). It has been observed that loss of *GLO^T* expression in short homostyle mutants resulted in loss of *KFB^T* in the petals (Figure 4.4). If the *GLO^T* transcription factor was responsible for activating expression of *KFB^T*, then – in addition to degrading the target protein from whichever pathway *KFB^T* modulates – *KFB^T* may also initiate degradation of *GLO^T* to cease its own transcription and complete a negative regulatory feedback loop. However, *GLO^T* could not be the sole transcription factor with the capacity to regulate *KFB^T* expression because *KFB^T* is also significantly expressed in the gynoeceium (Figures 4.3 & 4.4), where *GLO^T* expression is absent. Furthermore, *GLO^T* is absent in *A. thaliana* and transgenic *Arabidopsis* plants containing *GUS* under transcriptional regulation of the *KFB^T* promoter exhibited expression (Figure 4.8-2), demonstrating that the *KFB^T* promoter can be active in the absence of *GLO^T* expression.

The significance of this *KFB^T*-*GLO^T* interaction cannot be determined because *KFB^T* also showed binding activity with *PvGLO* and *PvDEF*. Although this may have suggested a lack of *KFB^T* target specificity, all three genes are B-function MADS-box genes and *PvGLO* shares 73 % sequence similarity with *GLO^T* at the amino acid level (Figure 6.2-2A). Furthermore, *PvDEF* displayed homology to the N-terminus of *GLO^T* (Figure 6.2-2B). It is therefore unclear whether *KFB^T* possesses promiscuous binding ability or has evolved affinity for closely related B-function MADS-box proteins. Further constructs could have been developed using genes unrelated to the S locus, such as *PvPlena* (Cook, 2002), or non-MADS-box proteins to identify non-interacting partners for *KFB^T* but it was decided to fully invest time into the protein library screens instead.

A

GLO ^T	MGRGKIEIKRIENSNNRQVTYSKRRNGI IKKAKEISLLCDAQVSLVIFANS GKMHDYCSP	60
GLO	MGRGKVEIKRIENS NIRQVTYSNRRNGILKKAKEISVLCDAQVSLIIFSSSGKMHDYCSP	60
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****	
GLO ^T	KTPLINILDAYQKQSGNRLWDAKHENLSNEIERVKKENDNMQIELRHLKGEDVQSLHHE	120
GLO	NSSLINILDAYQKQSGIRLWDARHENLSNEIERVKKENDNMQIELRYLKGEDIQSLHHE	120
	: : *****:*****:*****:*****:*****:*****:*****:*****:*****	
GLO ^T	LMSIESALENGLACVRQREMEIYRMARENFADKERVLE DENRSLTYQMHHLVMDIEGGEM	180
GLO	LMSIEDALENGLTRVRERQMEIYRMAKDNFADKERLLE DENKRLGYKFQQV-----	171
	*****:*****: * *:*****:*****:*****:*****: * *::: : :	
GLO ^T	ENGYNYQS QMPFSFRVQPIQP NLQERI	207
GLO	-----MDMQMPCSYRVQPLQP NLHDQF	193
	: * * * * :*****:*****: : : :	

B

DEF	MARGKIQIKKIENATNRQVTYSKRRNGLFKKAGELAVLCDAKVSIIIMLSGTNKIHEFHSP	60
GLO ^T	MGRGKIEIKRIENSNNRQVTYSKRRNGI IKKAKEISLLCDAQVSLVIFANS GKMHDYCSP	60
	*.*****: * *:*****:*****:*****:*****:*****:*****:*****:*****	
DEF	DHVTTKQLFDQYQQTMGIDLWSSHYEKMQEDLRKRKEINKNLRQEIRQRMGDSLSELDLN	120
GLO ^T	KTP-LINILDAYQKQSGNRLWDAKHENLSNEIERVKKENDNMQIELRHLKGEDVQSLHHE	119
	. : : * * *: * * * . : : * : : : : : : * : * . * : : * * : * : : . * . :	
DEF	KLQTLQDMEKSLEAIRGRKYVIGNQIETSRKKVRNGEEVHRTLLEFDAREEDPHYGL	180
GLO ^T	ELMSIESALENGLACVRQREMEIYRMARENFADKERVLE DENRSLTYQMHHLVMDIEGGE	179
	: * : : * . : * : . * . : * * : : * . * * * * : * : * : : . * . *	
DEF	VDNGGDYDSVIGYTNEGEPRMLSLRLQPNHHNLASGGGGNGLTTYALL	228
GLO ^T	MENGYNYQS QMPFS-----FRVQPIQP NLQERI-----	207
	: : * * : * : * : : : : * : * * : * * .	

* = perfect match ⚡ = highly similar properties • = lowly similar properties

Figure 6.2-2: Alignment of GLO^T (DQ381428.1) amino acid sequence against those of PvGLO (EF119212.1) and PvDEF (DQ381427.1). A) The PvGLO and GLO^T proteins share 73 % identity across their sequences and a further 15 % of amino acids had similar properties. B) The PvDEF and GLO^T proteins are particularly homologous at their N-terminus.

6.3 Yeast Two-Hybrid Screen of *Arabidopsis* and *Primula* Libraries

Yeast two-hybrid screens of both the *Primula* and *Arabidopsis* cDNA libraries were carried out to identify potential KFB^T partner proteins. The *Primula* library was the most relevant and therefore primary focus of these experiments but the *Arabidopsis* library was additionally screened due to its high quality, well-annotated genome and reliability in experiments by Dr Causier at the University of Leeds. Although *Arabidopsis* is an unrelated model system, it could still be informative if KFB^T targets a homologous housekeeping protein not unique to *Primula*.

Positive colonies were grown and underwent miniprep before their library fragment was amplified via PCR and sequenced (Chapter 2.20.3). To overcome silent variation between primary nucleotide sequences, BLASTx was used to identify amino acid sequences from the yeast two-hybrid positives. Results with an E-value greater than 0.01 were discarded. Seventeen candidate protein partners of KFB^T remained (Table 6.3.1). Only the GTP protein and Rubisco were from the *Primula* screen; all others came from the *Arabidopsis* library.

The myo-inositol-1-phosphate synthase gene (Table 6.3.1) emerged four times in one screen. This is involved in the production of all inositol-containing compounds by catalysing the conversion of D-glucose 6-phosphate to 1L-myo-inositol-1-phosphate (Geiger & Jin, 2006). The IQ-domain 24 protein also appeared twice and has a calmodulin-binding motif involved in cell-signalling (Fischer *et al.*, 2013). Rotamase cyclophilin 5 (Table 6.3.1) also has a role in signal transduction (Borderies *et al.*, 2003). Ribosomal proteins occurred three times across two different repeats of the yeast two-hybrid assay. However, these have elsewhere been found to bind so frequently that they hamper yeast two-hybrid experiments (Dowd *et al.*, 2000) and were therefore disregarded as false positives.

Overexpression of a Krueppel-like (Table 6.3.1) zinc finger protein in *Medicago truncatula* resulted in an unusually long style compared to anthers, which were dehydrated in the mutant (Frugier *et al.*, 2000). This led to inefficient pollination and lower production of pods that contained fewer seeds. It could be hypothesised that degradation of this Krueppel factor by KFB^T in the ubiquitinase pathway would lead to the short style observed in wildtype thrum flowers, although degradation of brassinosteroids by CYP^T has already been connected with style dimorphism in *Primula* (Huu *et al.*, 2016).

The heat shock, cold-regulated and chaperonin proteins from Table 6.3.1 are all involved in stress response pathways. The cytochrome P450 family member, CYP76C1, is also

preferentially expressed during the hypersensitive response to stress (Boachon *et al.*, 2015). The Developmental and Cell Death protein is an asparagine-rich protein that promotes cell death (Hoepflinger *et al.*, 2011). Orotidine 5'-phosphate decarboxylase is involved in the biosynthesis of uridine monophosphate (Radzicka & Wolfenden, 1995). Deoxy-D-xylulose 5-phosphate synthase is the main rate-determining enzyme in the production of plastidial isoprenoids for photosynthesis and development (Wright & Phillips, 2014). The Nodulin MtN3 family has a well-characterised role in carbohydrate transport and root nodule formation for nitrogen fixation (Gamas *et al.*, 1996).

Table 6.3-1: Potential KFBT-interacting proteins identified via yeast two-hybrid screens of *Primula* and *Arabidopsis* cDNA libraries.

Name	Accession	Library Screen
HSP20-like chaperone superfamily protein	NM_100614	
TCP-1/CPN60 chaperonin family protein	NM_102295	
Soybean gene regulated by cold-2 (SRC2)	NM_100778	
Early response to dehydration 1 (ERD1)	NM_124486	
Development and cell death domain protein (DCD)	NM_128851	
Nodulin MtN3 family protein	NM_101997	
DNA topoisomerase 1 alpha	NM_001203614	
IQ-domain 24	NM_120806	<i>Arabidopsis thaliana</i>
Krueppel-like factor	NM_121701	
Cytochrome P450 (CYP76C1)	NM_130119	
Orotidine 5-phosphate decarboxylase	NM_104905	
Deoxy-D-xylulose 5-phosphate synthase (DEF)	NM_117647	
Rotamase cyclophilin 5	NM_119653	
Myo-inositol-1-phosphate synthase (MIPS)	NM_120143	
Ribosomal protein 1	NM_103469	
GTP binding elongation factor Tu family protein	RZB57532.1	
Ribulose biphosphate carboxylase small chain 1A	XP_018848375.1	<i>Primula vulgaris</i>

6.4 Discussion

A number of self-incompatibility mechanisms have been characterised and shown to involve the influx of reactive oxygen species (ROS) that disrupt the Ca²⁺ gradient in pollen tubes and arrest their growth, sometimes via programmed cell death (Serrano *et al.*, 2015; Franklin-Tong *et al.*, 1997; Kenrick *et al.*, 1986). Self-incompatibility has been associated with cytoskeletal modifications too, such as rapid microtubule depolymerisation (Poulter *et al.*, 2008). Actin is also a very early target of self-incompatibility signals in *Papaver* (Staiger & Franklin-Tong, 2003). An extensive literature review found most candidate protein partners for KFB^T (Table 6.3.1) were associated with pollination, cytoskeletal modifications and stress response pathway components, such as calcium signalling. This uncovered a potential role for KFB^T in *Primula* self-incompatibility. Although the *Primula* self-incompatibility system is sporophytic (Dowrick, 1956), the underlying machinery is currently unknown and could be novel or may even have evolved to be mechanistically more similar to characteristics from gametophytic self-incompatibility systems (Lu *et al.*, 2018). It was therefore important that information from both classifications were researched.

According to The Arabidopsis Information Resource (TAIR; www.arabidopsis.org), the Krueppel-like factor identified from this yeast two-hybrid assay is expressed highly in the gynoecium, like KFB^T (Figures 4.3, 4.4 & 4.8-2). Overexpression of this protein resulted in unusually long styles compared to the anthers, which were dehydrated and thus led to inefficient pollination and fewer seed pods each containing a low number of seeds (Frugier *et al.*, 2000). Although gynoecium dimorphism is a major facet of floral heteromorphy, it is unlikely that KFB^T is responsible for the short style of thrum plants; CYP^T is believed to be responsible for this (Huu *et al.*, 2016). The knockout mutant also exhibited nitrogen deficiency and its gene expression was responsive to nodule development (Frugier *et al.*, 2000). Although nodulin was another candidate from this yeast two-hybrid assay (Table 6.3.1), these two proteins are unlikely to be linked.

The cytochrome P450 protein (CYP76C1; Table 6.3.1) is a major linalool metabolising enzyme expressed greatly and almost uniquely in *Arabidopsis* flower petals (Boachon *et al.*, 2015). Linalool compounds have a role in plant defence against insects (Müller *et al.*, 2009). This could be interpreted to suggest KFB^T temporarily inhibits production of this plant defence chemical to allow a window in which essential insects are attracted for cross-pollination. Although improbable, this proposition would be a completely novel and

previously overlooked role for the *Primula S* locus. A simple experiment to test this would be to assess whether thrum to pin crosses occur more often than pin to thrum. Furthermore, loss of *KFB^T* in short homostyle petal tissue (Figure 4.4) would lead to fewer pollinator visits to this mutant. This could also be tested by qPCR to assay *KFB^T* levels between day and night, as Darwin (1862) proposed that *Primula* are pollinated by moths at night.

The majority of yeast two-hybrid positives here appeared to be largely relevant to self-incompatibility. Nodulin was one of two top candidates for the pistil-component of self-incompatibility in *Senecio squalidus* and, like *Primula*, this species has a sporophytic self-incompatibility system controlled by a single *S* locus (Allen *et al.*, 2011). Myo-inositol oxygenase was another candidate mediator of self-incompatibility (Allen *et al.*, 2011), which is involved in the same pathway as myo-inositol-1-phosphate synthase (Loewus & Murthy, 2000) – another potential *KFB^T* partner protein (Table 6.3.1). The authors also identified an F-box protein as the top pollen-component of *S. squalidus* self-incompatibility and *KFB^T* is an F-box protein too.

The exact mechanism of Nodulin action in self-incompatibility was not defined but it has known roles within pollen development of *Arabidopsis thaliana* (Guan *et al.*, 2008) and *Oryza sativa* (Yang *et al.*, 2006). Its presence in stigmatic papillae cells infers a likely function in pollen recognition on the stigma by regulating traffic during pollen hydration and germination (Allen *et al.*, 2011). A nodulin 26-like intrinsic protein demonstrated heightened expression levels after incompatible self-pollination in *Fragaria viridis* (Du *et al.*, 2019). The *KFB^T*-mediated degradation of Nodulin in *Primula* may therefore breakdown self-incompatibility and permit pollination. Nodulin also has a role in the hypersensitive defence response (Gamas *et al.*, 1996). Furthermore, a nodulin 24-like gene was identified as a HT-family member (Kondo & McClure, 2008). Self-compatible *Solanum lycopersicum* have mutations in both *HT-A* and *HT-B* genes (Kondo *et al.*, 2002). HT proteins were first identified in *Nicotiana glauca* and are responsible for pollen rejection (McClure *et al.*, 1999). Antisense suppression of the *HT-B* pistil-expressed gene caused loss of gametophytic self-incompatibility in *Petunia* (Puerta *et al.*, 2009).

Both the overexpression or suppression of early nodulin-like (ENODL) genes in *Arabidopsis* compromised pollen tube reception (Hou *et al.*, 2016). This conversely suggests that degradation activity by *KFB^T* would therefore cause self-incompatibility instead of permitting compatibility, as proposed earlier. However, *KFB^T* is more similar to Nodulin

than ENODL (not shown). A nodulin-like gene is associated with gametophytic factor 1, which is a locus linked to gametophytic self-incompatibility in maize (Bloom, 2011). Nodulin has therefore been implicated in both sporophytic and gametophytic self-incompatibility systems and this supports the earlier statement that *Primula* investigations must be open to both modes of operation.

The second of the two self-incompatibility candidates from the *Senecio* investigation by Allen (*et al.*, 2011) was a pistil-specific membrane associated protein (MAP). Splicing factor 35 (SC35; Table 6.3.1) is also a candidate MAP (Derbyshire *et al.*, 2015), the depletion of which led to altered genes related to the transduction of plant hormone signals (Yan *et al.*, 2017). A Krueppel factor and orotidine-5'-phosphate decarboxylase emerged as *Primula* matches in the yeast two-hybrid (Table 6.3.2). These are both also associated with splicing. This role could be interesting as *GLO^T* is alternatively spliced in both thrums and long homostyles (Li *et al.*, unpublished). Only exon one of *GLO^T* has been detected in short homostyle transcriptomes (Li *et al.*, unpublished).

The SC35 protein has a role in transcriptional elongation and is associated with microtubules (Derbyshire *et al.*, 2015). Its knockout led to altered phyllotaxy (Yan *et al.*, 2017), which was also a mutant phenotype associated with the DNA topoisomerase 1 alpha candidate from Table 6.3.1 (Takahashi *et al.*, 2002). This was one of 100 proteins with a bioinformatically predicted relation to cell wall synthesis (Zhou *et al.*, 2010). Liquid chromatography-mass spectrometry found it was ubiquitinated (Manzano *et al.*, 2008) and this is at least potential evidence of its inclusion in the same pathway as KFB^T.

Two candidate heat shock and chaperonin proteins emerged from the yeast two-hybrid assay (Table 6.3.1). Expression of heat shock and chaperone proteins was shown to double during self-incompatible pollination of *Papaver rhoeas* (Poulter *et al.*, 2011). The manufacture of cytoskeleton-associated proteins also doubled and those involved with signal transduction increased by 50 % (Poulter *et al.*, 2011). Chaperonin 60 (CPN60) was specifically highlighted by the KFB^T screen (Table 6.3.1) and this protein was uniquely present in samples undergoing the self-incompatible reaction (Poulter *et al.*, 2011). It could therefore be interpreted that degradation of CPN60 via KFB^T activity would permit self-compatibility in *Primula*. Moreover, chaperonin 10 was downregulated in non-heading Chinese cabbage 15 mins after compatible pollination and upregulated in self-incompatible plants (Wang *et al.*, 2014). In *Solanum chacoense*, both a 68 kDa heat shock protein and a

60 kDa chaperonin protein were downregulated during self-incompatibility (Vyetrogon *et al.*, 2007).

The upregulation of heat shock, calcium binding and calmodulin proteins have all been observed in response to pollen germination and pollen tube growth (Wang *et al.*, 2008). Calcium signalling has also previously been linked to gametophytic self-incompatibility (Iwano *et al.*, 2015). IQ-domain 24 emerged from this yeast two-hybrid screen (Table 6.3.1) and is believed to be the site of action for Ca²⁺/calmodulin units (Petegem *et al.*, 2005). Calmodulin is a calcium sensor that regulates voltage-gated calcium channels (Spitzer, 2008). Evidence suggests the IQ-domain recruits calmodulin and tethers it to the channel for detection of nearby calcium (Bürstenbinder *et al.*, 2017). Degradation of this protein by KFB^T could therefore prevent the Ca²⁺ influx associated with self-incompatibility and permit compatibility instead. Interestingly, an IQ-domain protein (AT3G52870) was one of eleven gene candidates – alongside *CYP^T* – that demonstrated differential expression in *P. vulgaris* styles (Huu *et al.*, 2016). Although this candidate showed raised expression in the pin style (Huu *et al.*, 2016), mass differential expression analysis in *P. vulgaris* (Chapter 3.8) proposed that an IQM2-like IQ-domain containing protein could be potentially under-expressed in pin and short homostyle flowers (Figure 6.6) – which both lack a functional *GLO^T* gene (Li *et al.*, 2016).

Differential Expression Between *P. vulgaris* Flower Morphs of a Gene Encoding an IQM2-like Protein

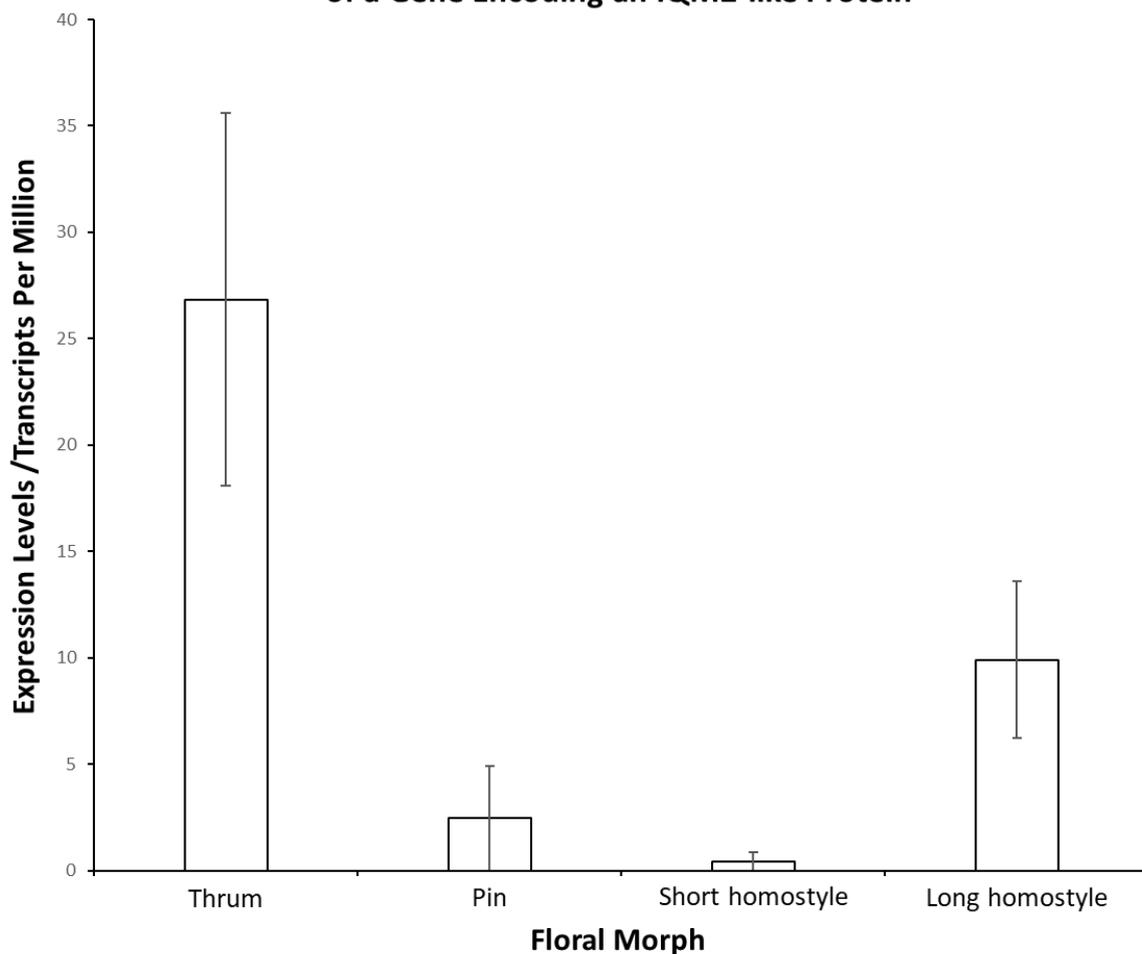


Figure 6.6: Differential expression of an IQM2-like gene between thrum, pin and homostyle flowers. This gene encodes an IQ domain-containing protein and emerged from an RNA-Seq screen to identify the most differentially expressed genes between morphs. An IQ domain-containing gene was also found to be upregulated in pin styles (Huu *et al.*, 2016) and a different IQ domain-containing candidate was identified in a yeast two-hybrid assay designed to investigate KFB^T-binding proteins.

The IQ-domain protein is also linked to dynamic organisation of microtubules (Bürstenbinder *et al.*, 2017). As the microtubule cytoskeleton coordinates direction of cell growth and expansion, initial conclusions alternatively linked this with suppression of *Primula* cell shape in the thrum upper corolla tube to support a hypothesis preliminarily raised by earlier qPCR experiments (Chapter 4.8). However, as ROS and actin filaments exhibit a known interaction with calmodulin during self-incompatible pollen tube growth (Jiang *et al.*, 2014), the self-incompatible hypothesis is favoured.

Binding of calmodulin to the *S* locus receptor kinase of *Brassica oleraceae* has been observed (Vanoosthuysse *et al.*, 2003). The Developmental and Cell Death protein (Table 6.3.1) was also found to interact with a calcium binding protein that shares sequence similarity with calmodulin (Hoepflinger *et al.*, 2011). Like heat shock proteins and

chaperonins, this protein has characterised involvement in stress pathways responsive to touch, ozone and salt (Hoepflinger *et al.*, 2011). Self-incompatibility has itself been considered a form of stress (Poulter *et al.*, 2011). Experiments with the *GUS* reporter gene suggested KFB^T expression could possibly be stress-responsive (Figure 4.8-2).

SRC2 is another Ca²⁺-responsive protein upregulated in response to stress, pollen germination and pollen tube growth (Wang *et al.*, 2008). It enhances ROS production (Kawarazaki *et al.*, 2013) and therefore inhibition by KFB^T may prevent arrest of pollen tube growth and permit compatibility. SRC2 is modulated by nitrous oxide and this is involved in ROS production, the hypersensitive response and programmed cell death, which are all related to self-incompatibility (Polverari *et al.*, 2003). It is also regulated by brassinosteroid (Tang *et al.*, 2008) and in *Primula* this hormone is modulated by CYP^T from the *S* locus to mediate style height dimorphism (Huu *et al.*, 2016).

Similarly to SRC2, ribulose biphosphate carboxylase small chain 1A (Rubisco) is regulated by cold temperatures (Kawamura & Uemura, 2003). It emerged as a positive from the KFB^T yeast two-hybrid screen and is downregulated in *Arabidopsis* GAPC-1 mutants that demonstrate defective pollen tube germination (Rius *et al.*, 2008). It was also downregulated in response to senescence in both wheat (Wittenbach, 1979) and rice (Makino *et al.*, 1984). ERD1 (Table 6.3.1) responds to senescence too (Montandon *et al.*, 2019). This is a member of the heat shock protein family discussed previously.

A third senescence-associated protein presented by this assay was the pheophytinase enzyme involved in chlorophyll breakdown (He *et al.*, 2018). No significant connection could be found with heteromorphy or self-incompatibility, except for senescence-induced expression of an S-RNase gene homologue in *Antirrhinum* (Liang *et al.*, 2002). S-RNases constitute the female self-incompatibility component in some systems (Bredemeijer & Blaas, 1981) and a self-incompatibility S1 gene from *Corchorus capsularis* was found to be differentially expressed between *P. vulgaris* morphs (Figure 3.8-3D). The GTP binding elongation factor Tu family protein candidate (Table 6.3.1) from the KFB^T yeast two-hybrid screen has been shown to bind with the S-RNase of *Solanum chacoense* and is accountable for the actin cytoskeleton disruption idiosyncratic of self-incompatibility (Soulard *et al.*, 2014). GTPases are a large protein family with well-documented roles in pollen tube growth and cytoskeletal modifications (McClure & Franklin-Tong, 2006).

In addition to nodulin, heat shock, chaperone, signal transduction, SRC2 and cytoskeletal proteins, myo-inositol-1-phosphate synthase also exhibited upregulation in response to pollen germination (Dai *et al.*, 2007). This enzyme is integral to the production of all inositol-containing compounds (Loewus & Murthy, 2000). Myo-inositol is taken up readily in germinating lily pollen for polysaccharide biosynthesis in the cell wall (Stanley & Loewus, 1964). Secondary wall thickening and callose plug formation during self-incompatible inhibition of pollination utilises such polysaccharides as pectin, callose and cellulose (Linskens, 2975).

Phytic acid – an inositol polyphosphate – is degraded more rapidly during incompatible than compatible pollen germination (Jackson *et al.*, 1983). This phytase activity was measured in *Petunia* and highest levels were detected in mature pollen (Jackson *et al.*, 1983). This could correspond with the significant spike in *KFB^T* transcription detected in thrum anthers during the pollen maturation stages of flower development (Figure 4.4). High quantities of phytic acid have been found in the pollen of species that require tube growth longer than ~4 mm (Jackson *et al.*, 1982). This may include thrum plants because wildtype pin styles are substantially longer than 4 mm.

Remarkable tolerances to drought, salinity and low temperature have been observed in *Mesembryanthemum crystallinum*, which accumulates high D-pinitol levels through methylation of a myo-inositol precursor (Agarie *et al.*, 2007). Silencing of myo-inositol-1-phosphate synthase in rice led to increased abscisic acid sensitivity and lower levels of myo-inositol, which usually protect the cell against reactive oxygen species (Ali *et al.*, 2013). Therefore, if *KFB^T* degraded this enzyme, the resultant decreased myo-inositol levels would leave cells more vulnerable to stress. This would potentially lead to cell death and the consequential arrest of pollen tube growth. A role for *KFB^T* in self-incompatibility is thereby inferred. Expression analysis with the *GUS* reporter gene potentially confirmed *KFB^T* transcription in an area expected to contain raised abscisic acid levels within the mature gynoecia of *Arabidopsis* flowers undergoing abscission (Figure 4.8-2).

Some of the candidate *KFB^T* protein partners (Table 6.3.1) were expected to be false positives and not all of them could be strongly linked to floral heteromorphy. Rotamase cyclophilin 5 is a cell-wall bound protein involved in signal transduction and protein refolding (Borderies *et al.*, 2003). No literature relevant to floral heteromorphy could be found for this protein or deoxy-D-xylulose 5-phosphate synthase, the latter of which is passed to heat shock proteins for either degradation or proper refolding (Pulido *et al.*,

2016). The orotidine 5'-phosphate decarboxylase candidate KFB^T partner (Table 6.3.1) has the largest rate enhancement of any known enzyme and catalyses the final step of uridine monophosphate biosynthesis (Radzicka & Wolfenden, 1995). Uridine phosphates span extensive roles in plants. For example, uridine diphosphate sugar pyrophosphorylase has a role in the self-incompatibility response of non-heading Chinese cabbage (Wang *et al.*, 2013a). It is downregulated in self-incompatible pistils, possibly leading to degradation of sucrose (Wang *et al.*, 2013b). Glucose deficiency is linked to male sterility (Meng *et al.*, 2009) and loss of function of this enzyme resulted in shrunken and collapsed pollen grains in *Arabidopsis* (Schnurr *et al.*, 2006). Transcription of a UDP-N-acetylglucosamine transferase gene was found to be significantly downregulated in short and long homostyle mutants (Figure 3.8-3E). This RNA-Seq data also presented a UDP-galactose/glucose transporter that exhibited no expression in pins or thrums, marginal transcription in short homostyles and significant overexpression in long homostyle mutants (Figure 3.8-3L).

Although numerous yeast two-hybrid candidates were raised in this assay, they may share a common KFB^T-binding site. No such site could be identified at the amino acid level however one may exist in the final 3-dimensional folded protein structures. One argument for KFB^T permitting compatibility in *Primula* (instead of establishing self-incompatibility) is that no phenotype was observed in mutant *Arabidopsis* flowers and normal seed set occurred; these plants are naturally self-compatible.

Conclusions

There are believed to be ~155 Kelch proteins in *Primula vulgaris* (Figure 3.2-2). The most similar to KFB^T shared ~45 % similarity (Table 3.2) and the most homologous non-*Primula* proteins were KMD1 and KMD2 from *Arabidopsis*, which both shared ~32 % similarity (Table 3.4). The KFB^T protein contains 5.5 Kelch repeats and an N-terminal F-box domain (Figure 3.1). A transcription start site has been predicted ~900 bp upstream of the KFB^T start codon (Figure 3.6.3-3).

The presence of KFB^T has been confirmed in fifteen other *Primulaceae* species (Figure 3.3.4). The Kelch domain is highly conserved between these species, with a variable region in between the fourth and fifth Kelch repeats and increasing variation toward the C-terminus of the F-box domain (Figure 3.3.5-3). There is further purifying selection across the C-terminus of the KFB^T protein that does not correspond with an identified Kelch repeat (Figure 3.3.5-3), which supports the suggestion elsewhere that a C-terminal structure may cap the β -barrel formed by Kelch propellers (Li *et al.*, 2004). This was also observed when KFB^T was aligned to 154 *P. vulgaris* Kelch proteins and the four *Arabidopsis* KMD homologues (Figure 3.5).

Greatest expression levels of KFB^T were detected in the *P. vulgaris* style via qPCR, with increasing transcription in the anthers as the flower approached maturity (Figure 4.4). Activity in these reproductive organs inferred a potential role for KFB^T in the heteromorphic self-incompatibility system. Gynoecium expression was reinforced by a *GUS* reporter gene assay (Figure 4.8-2). This self-incompatibility hypothesis was supported by positives from a yeast two-hybrid screen, of which almost all of them could be linked to self-incompatibility. It is currently unclear whether KFB^T acts to permit pollination or instigate incompatibility. No KFB^T , CCM^T or PUM^T transcripts were detected in the self-compatible *P. oreodoxa* (Zhao *et al.*, 2019). A reduction of KFB^T expression within self-compatible short homostyle petals was also observed (Figure 4.3), however activity in petals would not be expected to effect incompatibility. Alternatively, KFB^T transcription appeared to increase in the styles of self-compatible long homostyle flowers (Figure 4.3).

One possible model is that *Primula* are natively unreceptive to all pollen and KFB^T in thrum pollen breaks down this barrier to permit pollination upon crossing to pin. Similarly, KFB^T in the thrum style would break the barrier of incompatibility to permit crossing from pin pollen. However, during thrum self-fertilisation, KFB^T in the pollen may homodimerize with

stylar KFB^T . This homodimerization has been proven by yeast two-hybrid (Figure 6.2-1). Such a protein-protein interaction may disable KFB^T functionality and maintain the native incompatible state of the thrum plant. However, it is currently unexplained why KFB^T would not always homodimerize with neighbouring KFB^T molecules in the same tissue.

This model depends on a balance between KFB^T in the pollen and style to limit the amount of free KFB^T to breakdown self-incompatibility and permit fertilisation. The significantly increased quantity of KFB^T transcription in long homostyle gynoecia (Figure 4.3) could be enough to establish an imbalance that renders sufficient KFB^T proteins free to cause the self-compatibility phenotype consistently observed in these mutants.

The gynoecium:anther KFB^T ratios in thrum, short homostyle and long homostyle flowers are 4.38, 1.36 and 9.66 respectively (Figure 4.3). This suggests it could be the relative increase of KFB^T expression in short homostyle anthers that causes their self-compatibility phenotype. However, these figures are taken from pools containing flowers at various stages of development and not only at the phase of anthesis, therefore KFB^T may be overrepresented in the gynoecium. This is because KFB^T transcription begins earlier here than anthers but nothing is known regarding the post-translational persistence of KFB^T prior to undergoing protein turnover.

The existence of two separate pin and thrum self-incompatibility systems has previously been proposed (Lu *et al.*, 2018), though no genomic differences outside of the *S* locus have yet been found (Li *et al.*, 2016; Cocker *et al.*, 2018). This novel model potentially explains how one protein could manage self-incompatibility systems in both thrum and pin plants; the latter of which does not possess the *S* locus (Li *et al.*, 2016). Targeted knockout of KFB^T in *P. vulgaris* by way of CRISPR would be the ideal method for confirming the function of this gene in the future. The KFB^T coding sequence and promoter (Figure 5.3) should also be transformed into the pin genome, to observe how it behaves in a *Primula* plant that does not natively contain the *S* locus.

Appendix A

Bespoke scripts written for this project are included here. Red text indicates code. Green text indicates input files required from the user. Comments in black text alongside a hash symbol are notes provided to aid the user.

A1 Obtain coding sequences from a genome using a GFF annotation file

A script was written that obtains all coding sequences from an annotated genome. The genome should be provided in FASTA format and the annotation should be in General Feature Format (GFF). It will produce a single FASTA output file containing each gene ID and the corresponding coding sequence. Sequences that do not begin and end correctly with start and stop codons are dismissed.

```
# Initialise variable flags for later
skip=0
gene=0

# Scan gff file line-by-line
cat Primula.gff | while read line
# Genes in the gff are separated by a row of '###'. If we haven't yet encountered a new gene...
do if [ "$line" != "###" ]
# If the 'skip' flag has not been raised, the gff line refers to an mRNA other than the first variant...
then if [ $skip -eq 0 ] && [ $(echo $line | awk '{print $3}') == "mRNA" ] && [ $(echo $line | awk -F ';' '{print $(NF-1)}' |
awk -F ':' '{print $2}') -gt 1 ]
# Then trigger the skip flag
then skip=1
# If the 'skip' flag has not been raised and the gff line refers to a coding sequence
elif [ $skip -eq 0 ] && [ $(echo $line | awk '{print $3}') == "CDS" ]
# Then print the gene ID, alongside the scaffold number and the start and stop positions
then echo -n "$gene "
echo $line | awk '{print $1, $4, $5}'
# If the 'skip' flag has not been raised and the gff line refers to a gene, save the gene ID from the final column
elif [ $skip -eq 0 ] && [ $(echo $line | awk '{print $3}') == "gene" ]
then gene=$(echo $line | awk -F 'ID=' '{print $2}' | awk -F ';' '{print $1}')
fi
# Otherwise, you must have encountered a new gene. Reset the 'skip' flag and proceed.
else
skip=0
fi
done | while read line
# Go through the output, echo the gene ID with the corresponding sequence from the P. vulgaris genome
do echo $(echo $line | awk '{print $1}')$(grep -w -m1 -A1 $(echo $line | awk '{print $2}') ../pvulgaris_LH_v2.fa | tail -n 1 |
fold -w1 | awk -v start=$(echo $line | awk '{print $3}') -v end=$(echo $line | awk '{print $4}') 'NR>=start && NR<=end' | tr -d '\n')
done > Primula.CDS_chunks

# For each gene...
for item in $(awk '{print $1}' Primula.CDS_chunks | sort | uniq)
# Find the coding sequence chunks and paste them together into a coherent sequence
do sequence=$(grep -w $item CDS_chunks | awk '{print $2}' | tr -d '\n')
# Check the sequence begins with a start codon
if [ $(echo $sequence | fold -w3 | head -n 1) == "ATG" ]
# Check the sequence ends with a stop codon
then if [ $(echo $sequence | fold -w3 | tail -n 1) == "TAG" ] || [ $(echo $sequence | fold -w3 | tail -n 1) == "TAA" ] || [
$(echo $sequence | fold -w3 | tail -n 1) == "TGA" ]
# If so, print the gene ID and sequence in FASTA format
then echo '>$item
echo $sequence
fi
done
rm Primula.CDS_chunks
```

A2 Score Alignments to Calculate Similarity Across Sequences

A script was written to interpret Clustal alignments for scoring similarity across sequences. This was useful for calculating percentage similarity and visualising conserved regions across multiple sequence samples via line graphs. The user must provide a FASTA file containing the template sequence and the filename must end in 'template.fa'. At least one query sequence must also be provided in FASTA format, ending with 'query.fa'. Multiple query sequences may be provided but they must each be kept in a different file. Version 2.1 of the ClustalW software is required (Larkin *et al.*, 2007). A 'Percentage_Similarities.txt' file is produced containing the similarity calculation for every query sequence included. A table containing hits alongside their corresponding base or amino acid position is also produced, suitable for plotting on a line graph.

```

# Blank files are prepared for later in case temporary files remain from previous runs.
echo -n > clustal.info
echo -n > results.temp
echo -n > Similarity.forR

# Queries are kept as fasta files ending in 'query.fa' and a useful list of these query names is stored in a variable.
query_list=$(ls *query.fa | sed 's/.query.fa//g')

# Separate files containing the template and query sequences are made.
for file in *query.fa
do cat $file > "$file".forClustal
cat *.template.fa >> "$file".forClustal
done

# Alignments are carried out between template and query sequences. Output information is saved into 'clustal.info'
source clustalw-2.1
for file in *.forClustal
do clustalw $file >> clustal.info
done

# Each Clustal alignment file has the aligned query sequences separated from the template sequences.
for query in $query_list
do grep $query "$query".query.fa.aln | awk '{print $2}' | tr -d '\n' | fold -w1 > "$query".aligned
# These are subsequently made vertical....
grep $(ls *.template.fa | awk -F '.' {print $1}') "$query".query.fa.aln | awk '{print $2}' | tr -d '\n' |
fold -w1 > "$query".template
# .... and pasted back together and matches or misalignments are presented as 1 or 0
paste "$query".template "$query".aligned | awk '$1!="-" {if ($1==$2) print $0, 1; else print $0, 0}' >
"$query".final_alignment
done

# The percentage similarity is calculated between each query and template sequence.
for query in $query_list
do awk -v query=$query '{sum+= $3} END{print query, ((sum/NR)*100)}' "$query".final_alignment
awk '$3==1 {print FNR}' "$query".final_alignment >> results.temp
done | sort -hrk2 > Percentage_Similarities.txt

# The percentage of queries in which each base appears is printed alongside the base position, for line graph plotting
number_of_queries=$(ls *query.fa | wc -l)
template_length=$(grep -A1 '>' *template.fa | tail -n 1 | wc -c)
base=1
while [ $base -le $template_length ]
do echo $base $(grep -wc $base results.temp | awk -v number_of_queries=$number_of_queries '{printf
"%0f\n", ($1/number_of_queries)*100}')
base=$((base+1))
done >> Similarity.forR

# All temporary files are removed.
rm clustal.info results.temp *forClustal *aligned *alignment *dnd *aln *template

```

A3 Generate Consensus Sequence from Multiple Nucleotide Sequences

A script was written to generate a consensus sequence from multiple nucleotide sequences. This was used for subsequently enumerating synonymous and non-synonymous substitutions across various species, in an investigation inspired by K_a/K_s statistics. Queries should be provided by the user in separate files as vertical sequences occupying a single column, with one base per line. The 'fold -w1' command will convert horizontal sequences into the correct format, if any surplus text such as headers are removed first. The script generates a single consensus nucleotide sequence.

```
# The vertical query sequences were pasted together into columns
paste *vertical_query | while read line
    # Each nucleotide was quantified across the sequences and the most common was output into the consensus
    do for base in A T G C
        do echo $base
            echo $line | fold -w1 | grep -c $base
        done | paste - - | sort -hk2 | tail -n 1 | awk '{print $1}'
    done | tr -d '\n' > consensus_nucleotide
# A newline character was added to the very end of the consensus sequence file.
echo "" >> consensus_nucleotide
```

A4 Count Synonymous and Non-Synonymous Substitutions

This script was written to detect mutations between query and consensus nucleotide sequences and subsequently determine whether they represented synonymous or non-synonymous substitutions. The queries should be provided in separate files as vertical sequences occupying a single column, with one base per line. The 'fold -w1' command will convert horizontal sequences into the correct format, if any surplus text such as headers are removed first. The consensus nucleotide sequence should be provided in a file, with no headers or surplus text. A codon table is also required, containing each nucleotide triplet alongside its corresponding single letter amino acid symbol in an adjacent column. The script will output a table containing each codon position alongside the number of synonymous and non-synonymous substitutions identified therein.

```
# Variables are initialised for later.
codonnumber=1
synonymous_substitutions=0
non_synonymous=0
for file in *vertical_query
do
  # Prepare output file.
  do echo -n > "$file".mutation_count
  # Take each codon from query
  cat $file | tr -d '\n' | fold -w3 | while read codon
  do
    # Find corresponding codon in consensus sequence
    do comparecodon=$(cat consensus_nucleotide | fold -w3 | awk NR==$codonnumber)
    # If codon differs....
    if [ "$codon" != "$comparecodon" ]
    then
      # Take each base in codon
      then for base in 1 2 3
      do if [ $(echo $codon | fold -w1 | awk NR==$base) == '.' ]
      then continue
      # If base differs from consensus....
      elif [ $(echo $codon | fold -w1 | awk NR==$base) != $(echo $comparecodon | fold -w1 | awk NR==$base) ]
      then testcodon=$(for test in 1 2 3
      do if [ $base -eq $test ]
      then echo -n $codon | fold -w1 | awk NR==$test
      else echo -n $comparecodon | fold -w1 | awk NR==$test
      fi
      done)
      testcodon=$(echo $testcodon | sed 's//g')
      # Check if differing base changes encoded amino acid
      if [ $(grep $testcodon codon_table | awk '{print $2}') == $(grep $comparecodon codon_table | awk '{print $2}') ]
      # If not, increase the synonymous substitution count for this codon
      then synonymous_substitutions=$((synonymous_substitutions+1))
      else
      # If yes, increase the non-synonymous substitution count for this codon
      non_synonymous=$((non_synonymous+1))
      fi
    fi
  done
fi
# Print the scores alongside the current codon number into a log file for this sample
echo $codonnumber $synonymous_substitutions $non_synonymous >> "$file".mutation_count
# Move codon number forward and reinitialise variables
codonnumber=$((codonnumber+1))
synonymous_substitutions=0
non_synonymous=0
done
# Reinitialise codon number for next file
codonnumber=1
done
```

A5 Dot Matrix Alignment to Identify Self-Complementary Regions

This script will accept a user sequence and carry out a dot matrix alignment against its reverse complement to identify potentially self-complementary regions within it. The user should provide both forward and reverse complement sequences transformed vertically into a single column, so one base occupies one line. The user can select a kmer size to use in the process by editing the top line of the script. Matching forward and reverse positions are printed, suitable for plotting on a scatter graph for analysis.

```
# Variable for kmer size initialised. This should be selected by the user, or even ran in a loop to test many.
kmer_size=14

# Empty files are prepared for later.
echo -n " " > forward."$kmer_size"mer
echo -n " " > reverse."$kmer_size"mer

# User sequences are broken into kmers.
begin=1
end=$kmer_size
while [ $end -le $(cat forward.folded | wc -l) ]
do echo $(awk -v begin=$begin -v end=$end 'NR>=begin&&NR<=end' forward.folded | tr -d '\n') >>
forward."$kmer_size"mer
echo $(awk -v begin=$begin -v end=$end 'NR>=begin&&NR<=end' reverse.folded | tr -d '\n') >> reverse."$kmer_size"mer
begin=$((begin+1))
end=$((end+1))
done

# Matching kmers shared by the forward and reverse complement sequences are identified.
fnumber=1
rnumber=1
cat forward."$kmer_size"mer | while read fline
do cat reverse."$kmer_size"mer | while read rline
do if [ $fline == $rline ]
then # The matching forward and reverse kmer numbers (positions) are printed.
echo $fnumber $rnumber
fi
rnumber=$((rnumber+1))
done
fnumber=$((fnumber+1))
done > rc."$kmer_size"mer.results
```

A6 Identify Stem and Loop Structures in a Nucleotide Sequence

This script was used to identify potential stem and loop structures in the *KFB^T* nucleotide sequence. The input sequence was provided in a file without a header. Stems of five to fourteen with loops between three and twenty nucleotides were screened. The stem and loop structures were printed to screen alongside their lengths.

```
# Stem lengths between 5 and 14 bp are tested
for length in {5..14}
  # Each stem in the forward sequence is obtained
  do fold -w $length kf.fai | while read kmer
    # The reverse complement stem is obtained
    do reverse_kmer=$(
      line=$length
      while [ $line -gt 0 ]
        do echo $kmer | sed 'y/agtc/tcag/' | fold -w1 | awk NR==$line | tr -d '\n'
          line=$((line-1))
        done)
    # The sequence is queried for these stems separated by linkers of 3 to 20 nucleotides
    for linker in {3..20}
      do echo -n $kmer$(line=$linker; while [ $line -gt 0 ]; do echo -n '.'; line=$((line-1)); done)$reverse_kmer
        kmer_length: $length linker_length: $linker" "
          grep -c -E "$kmer.{linker}$reverse_kmer" kf.fai
        done
      done
    done | awk '$NF==1'
```

A7 Calculate Average Transcripts Per Million for Every Gene

This script was written to calculate the average TPM for every *Primula vulgaris* gene by using RNA-Seq data processed by StringTie (Pertea *et al.*, 2016). The user supplies the StringTie output file and a table containing each gene with the start and stop positions, which should easily be obtained from a GFF genome annotation file. The script lists each gene with its corresponding average expression level.

```
# User supplies a StringTie output file
StringTieFile=StringTie.out

# For each gene in the gff annotation...
cat ../genome_files/adjusted_gff | while read line
do # Print the scaffold name
  scaffold=$(echo $line | awk '{print $1}')
  # Print the start position
  begin=$(echo $line | awk '{print $4}')
  # Print the end position
  end=$(echo $line | awk '{print $5}')
  # Calculate the mean transcript depth across this region for each gene
  echo ID=$(echo $line | awk -F 'ID=' '{print $2}' | awk -F ';' '{print $1}') $(awk -v scaffold=$scaffold
'$1==scaffold&&$3=="transcript"' $StringTieFile | awk -v begin=$begin -v end=$end 'begin>$4&&begin<$5
|| end>$4&&end<$5
|| begin<$4&&end>$5' | awk -F "" '{sum+=$10} END{print sum/NR}')
done
```

A8 Find Closest Protein Matches in *Primula vulgaris*

This script took nucleotide sequences from yeast two-hybrid positives of an *Arabidopsis* library screen and found their closest match in the *P. vulgaris* genome. My CRanslate tool from GitHub (github.com/calumraine/cranslate) was used to translate the nucleotide sequences and version 2.1 of ClustalW (Larkin *et al.*, 2007) carried out the subsequent alignments. Multiple query sequences are permitted, each in separate FASTA files. A second FASTA file containing all amino acid sequences from *P. vulgaris* was also required – this was obtained via the script in Appendix A1.

```
# A variable is created to hold the filename so the script can easily be used with any file by only changing the top line.
file=user_file.fa

# Version 2.1 of the ClustalW multiple sequence alignment tool is used.
source clustalw-2.1

# An empty log file is setup and some variables are initialised for later.
echo > "$file".log
current_hits=0
top_hits=0
top_protein=NA

# The input file should be in fasta format, allowing the the length of the query sequence to be calculated.
length=$((($(awk NR==2 "$file" | wc -c)-1)))

# My CRanslate tool is used to translate the forward three frames of the nucleotide sequence (github.com/calumraine/cranslate)
cat "$file" | cranslate -F | paste -- | while read frame
# The genes.translated.fa file contains amino acid sequences translated from every gene in the P. vulgaris genome.
do cat genes.translated.fa | paste -- | while read protein
# Each amino acid sequence is individually added to a file containing the query sequence.
do echo "$frame" | awk '{print $1"\n"$2}' > "$file".current_query
echo "$protein" | awk '{print $1"\n"$2}' >> "$file".current_query
# These are subsequently aligned via Clustal.
clustalw "$file".current_query >>/dev/null
# The number of asterisks are enumerated and multiplied by three to score perfect matches
matches=$(cat "$file".aln | fold -w1 | grep -c '*')
# The number of colons are counted and doubled to score highly similar matches .
highly=$(awk 'NR>1 {print $2}' "$file".aln | fold -w1 | grep -c ':')
# The alignment is searched for periods to enumerate the number of amino acids sharing lowly similar properties
lowly=$(awk 'NR>1 {print $2}' "$file".aln | fold -w1 | grep -c '\.')
current_hits=$((matches*3+highly*2+lowly))
# The current highest score is held in a variable for final output. This means the log file below can be removed if desired.
if [ $current_hits -gt $top_hits ]
then top_hits=$current_hits
top_protein=$(echo $protein | awk '{print $1}' | sed 's/>/')
fi
# The current score is logged alongside a percentage of the maximum possible score.
echo $file $(echo $protein | awk '{print $1}' | sed 's/>/') $current_hits $(((current_hits*100)/length)) >> "$file".log
# Temporary alignment files are removed.
rm "$file".dnd "$file".aln
done
done

Upon completion, the log file can simply be sorted (sort -hrk3) to identify the protein with the best match. The log file was included for potential troubleshooting and to see other matches beyond the best hit. This script can be run in parallel with multiple queries, as long as the filenames are different.
```

Appendix B

A list of reagents used throughout the methods chapter is here included.

Product	Supplier	Catalogue Number
-Leu Dropout Supplement	Clontech	630414
-Leu/-Trp Dropout Supplement	Clontech	630417
-Leu/-Trp/-His Dropout Supplement	Clontech	630419
-Trp Dropout Supplement	Clontech	630413
BioMax MR Scientific Imaging Film	KODAK	8912560
Deoxyribonucleic acid, single stranded from salmon testes	Sigma-Aldrich	D7656
Dneasy® Plant Mini Kit	QIAGEN	69104
dNTP Set	Invitrogen™/ThermoFisher Scientific	10297018
Extract-N-Amp™ Plant PCR Kit	SIGMA	XNAP2
Gamborg B5 Medium Macro Salt Mixture	Duchefa Biochemie	M0304
Gateway™ LR Clonase™ II Enzyme Mix	Invitrogen™/ThermoFisher Scientific	11791020
Glass beads, acid washed	SIGMA	G8772
Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488)	abcam	ab150077
GoTaq® G2 Flexi DNA Polymerase	Promega	M7801
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems™/ThermoFisher Scientific	4368814
InstantBlue™ Ultrafast Protein Stain	Sigma-Aldrich	ISB1L
MAGNA Nylon Transfer Membrane	Osmonics Inc.	N00HG08250
Murashige & Skoog Medium Including Vitamins	Formedium	MS0150
Murashige & Skoog Medium Micro-Salt Mixture MS	Duchefa Biochemie	M0301
Murashige & Skoog Vitamin Mixture	Duchefa Biochemie	M0409
NuPAGE™ Antioxidant	Invitrogen™/ThermoFisher Scientific	NP0005
NuPAGE™ MOPS SDS Running Buffer (20x)	Invitrogen™/ThermoFisher Scientific	NP0001
NuPAGE™ Transfer Buffer (20x)	Invitrogen™/ThermoFisher Scientific	NP0006
One Shot™ ccdB Survival™ 2 T1R Competent Cells	Invitrogen™/ThermoFisher Scientific	A10460
One Shot™ TOP10 Chemically Competent E. coli	Invitrogen™/ThermoFisher Scientific	C404010
pCR™8/GW/TOPO™ Cloning Kit with One Shot™ Top10 E. coli	Invitrogen™/ThermoFisher Scientific	K250020
Phusion® High Fidelity DNA Polymerase	New England Biolabs	M05305
Plant RNA Isolation Aid	Ambion/Invitrogen™	AM9690
QIAquick® Gel Extraction Kit	QIAGEN	28704
QIAquick® PCR Purification Kit	QIAGEN	28104
Qubit™ dsDNA BR Assay Kit	Invitrogen™/ThermoFisher Scientific	Q32850
Qubit™ ssDNA Assay Kit	Invitrogen™/ThermoFisher Scientific	Q10212
Qubit™ RNA BR Assay Kit	Invitrogen™/ThermoFisher Scientific	Q10210
RNAqueous™ Total RNA Isolation Kit	Ambion/ThermoFisher Scientific	AM1912
RQ1 Rnase-Free Dnase	Promega	M6101
Silwet L-77	DE SANGOSSE	640
Subcloning Efficiency™ DH5α Competent Cells	Invitrogen™/ThermoFisher Scientific	18265017
SuperSignal™ West Femto Trial Kit	Thermo Scientific	34094
SYBR® Green JumpStart™ Taq Polymerase ReadyMix™	Sigma-Aldrich	S4438
Tissue-Tek® O.C.T. Compound	Sakura® Finetek/VWR	25608-930
VWR™ Pellet Mixer	VWR International	47747-370
Wizard® Plus SV Minipreps DNA Purification Systems	Promega	A1330

References

1. Adams, J., Kelso, R., Cooley, L., 2000, The kelch repeat superfamily of proteins: propellers of cell function, *Trends in Cell Biology*, 10(1):17-24
2. Agarie, S., Shimoda, T., Shimizu, Y., Baumann, K., Sunagawa, H., Kondo, A., Ueno, O., Nakahara, T., Nose, A., Cushman, J. 2007, Salt tolerance, salt accumulation, and ionic homeostasis in an epidermal bladder-cell-less mutant of the common ice plant *Mesembryanthemum crystallinum*, *Journal of Experimental Botany*, 58(8):1957-67
3. Ai, Y., Kron, E., Kao, T., 1991, S-alleles are retained and expressed in a self-compatible cultivar of *Petunia hybrida*, *Molecular and General Genetics*, 230:353-8
4. Ali, N., Paul, S., Gayen, D., Sarkar, S., Datta, S., Datta, K., 2013, RNAi mediated down regulation of myo-inositol-3-phosphate synthase to generate low phytate rice, *Rice*, 6(12)
5. Allen, A., Thorogood, C., Hegarty, M., Lexer, C., Hiscock, S., 2011, Pollen-pistil interactions and self-incompatibility in the Asteraceae: new insights from studies of *Senecio squalidus* (Oxford ragwort), *Annals of Botany*, 108(4):687-98
6. Alwen, A., Moreno, R., Vicente, O., Heberle-Bors, E., 1992, Plant endogenous β -glucuronidase activity: how to avoid interference with the use of the *E. coli* β -glucuronidase as a reporter gene in transgenic plants, *Transgenic Research*, 1(2):63-70
7. Al Wadi, H., Richards, A., 1993, Primary Homostyly in *Primula* L. Subgenus *Sphodylia* (Duby) Rupr. And the Evolution of Distyly in *Primula*, *New Phytologist*, 124(2):329-38
8. Ambion, 2008, RNAqueous® Kit, Invitrogen, Publication Number: 1912M Revision C
9. Andrade, M., Gonzalez-Guzman, M., Serrano, R., Rodriguez, P., 2001, A combination of the F-box motif and kelch repeats defines a large *Arabidopsis* family of F-box proteins, *Plant Molecular Biology*, 46:603-14
10. Araujo, R., Yoon, K., Ko, D., Smith, A., Qiao, M., Suresh, U., Burns, S., Penalva, L., 2012, Before It Gets Started: Regulating Translation at the 5' UTR, *Comparative and functional genomics*: 12:475731
11. Arunkumar, R., Wang, W., Wright, S., Barrett, S., 2017, The genetic architecture of tristily and its breakdown to self-fertilisation, *Molecular Ecology*, 26:752-65
12. Asakura, Y., Barkan, A., 2007, A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts, 19(12):3864-75
13. Athanasiou, A., Khosravi, D., Tamari, F., Shore, J., 2003, Characteriation and Localization of Short-Specific Polygalacturonase in Distylous *Turnera subulate* (*Turneraceae*), *American Journal of Botany*, 90(5):675-82
14. Athanasiou, A., Shore, J., 1997, Morph-Specific Proteins in Pollen and Styles of Distylous *Turnera* (*Turneraceae*), *Genetics*, 146(2):669-79
15. Babendure, J., Babendure, J., Ding, J., Tsien, R., 2006, Control of mammalian translation by mRNA structure near caps, *RNA*, 12(5):851-61
16. Bachmair, A., Novatchkova, M., Potuschak, T., Eisenhaber, F., 2001, Ubiquitylation in plants: a post-genomic look at a post-translational modification, *Trends in Plant Science*, 6(10):463-70
17. Bai, C., Sen, P., Hofmann, K., Ma, L., Goebel, M., Harper, J., Elledge, S., 1996, SKP1 Connects Cell Cycle Regulators to the Ubiquitin Proteolysis Machinery through a Novel Motif, the F-Box, *Cell*, 86(2):263-74
18. Barkan, A., Small, I., 2014, Pentatricopeptide Repeat Proteins in Plants, *Annual Review of Plant Biology*, 65:415-42
19. Bateson, W., Gregory, R., 1905, On the inheritance of heterostylism in *Primula*, *Proceedings of the Royal Society B: Biological Sciences*, 76(513):581-6
20. Baulcombe, D., 1999, Fast forward genetics based on virus-induced gene silencing, *Current Opinion in Plant Biology*, 2(2):109-13
21. Bernard, P., Couturier, M., 1992, Cell Killing by the F Plasmid CcdB Protein Involves Poisoning of DNA-Topoisomerase II Complexes, *Journal of Molecular Biology*, 226:735-45
22. Bernatzky, R., Galven, R., Rivers, B., 1995, S-related protein can be recombined with self-compatibility in interspecific derivatives of *Lycopersicon*, *Biochemical Genetics*, 33:215-25
23. Bertani, G., 1951, Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*, *Journal of Bacteriology*, 62(3):293-300
24. Bloom, J., 2011, Genomic localisation of the maize cross-incompatibility gene, Gametophyte factor 1 (*ga1*), *Maydica*, 56(4):379-87
25. Boachon, B., Junker, R., Miesch, L., Bassard, J., Höfer, R., Caillieudeaux, R., Seidel, D., Lesot, A., Heinrich, C., Ginglinger, J., Allouche, L., Vincent, B., Wahyuni, D., Paetz, C., Beran, F., Miesch, M.,

- Schneider, B., Leiss, K., Werck-Reinhart, D., 2015, CYP76C1 (Cytochrome P450)-Mediated Linalool Metabolism and the Formation of Volatile and Soluble Linalool Oxides in Arabidopsis Flowers: A Strategy for Defense against Floral Antagonists, *The Plant Cell*, 27(2972-90)
26. Bodmer, W., 1958, Natural crossing between homostyle plants of *Primula vulgaris*, *Heredity*, 12:363-70
 27. Bodmer, W., 1960, Genetics of homostyly in populations of *Primula vulgaris*, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 242(696):517-49
 28. Bodmer, W., 1984, Sex and generations of primroses, *Nature*, 310, 731
 29. Bolognesi, B., Lehner, B., 2016, Protein Overexpression: Reaching the limit, *eLife*, 7:e39804
 30. Borderies, G., Jamet, E., Lafitte, C., Rossignol, M., Jauneau, A., Boudart, G., Monsarrat, B., Esquerré - Tugayé, M., Boudet, A., Pont-Lezica, R., 2003, Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis, *Electrophoresis*, 24:3421-32
 31. Bork, P., Doolittle, R., 1994, *Drosophila kelch* motif is derived from a common enzyme fold, *Journal of Molecular Biology*, 236(5):1277-82
 32. Bower, M., Matias, D., Fernandes-Carvalho, E., Mazzurco, M., Gu, T., Rothstein, S., Goring, D., 1996, Two members of the thioredoxin-h family interact with the kinase domain of a Brassica S locus receptor kinase, *The Plant Cell*, 8:1641-50
 33. Bredemeijer, G., Blaas, J., 1981, S-specific proteins in styles of self-incompatible *Nicotiana glauca*, *Theoretical and Applied Genetics*, 59:185-90
 34. Bucher, P., 1990, Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences, *Journal of Molecular Biology*, 212:563-78
 35. Burke, T., Kadonaga, J., 1996, *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters, *Genes & Development*, 10(6):711-24
 36. Burrows, B., McCubbin, A., 2017, Sequencing the genomic regions flanking S-linked *PvGLO* sequences confirms the presence of two *GLO* loci, one of which lies adjacent to the style-length determinant gene *CYP734A50*, *Plant Reproduction*, 30(1):53-67
 37. Burrows, B., McCubbin, A., 2018, Examination of S-Locus Regulated Differential Expression in *Primula vulgaris* Floral Development, *Plants*, 7(2):38
 38. Bürstenbinder, K., Mitra, D., Quegwer, J., 2017, Functions of IQD proteins as hubs in cellular calcium and auxin signalling: A toolbox for shape formation and tissue-specification in plants?, *Plant Signaling & Behaviour*, 12(6):e1331198
 39. Cabrillac, D., Cock, J., Dumas, C., Gaude, T., 2001, The S-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins, *Nature*, 410:220-3
 40. Catic, A., Ploegh, H., 2005, Ubiquitin – conserved protein or selfish gene?, *Trends in Biochemical Sciences*, 30(11):600-4
 41. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., Prasher, D., 1994, Green fluorescent protein as a marker for gene expression, *Science*, 263(5148):802-5
 42. Chapman, A., 2009, Numbers of Living Species in Australia and the World, Australian Biodiversity Information Services, 2nd edition
 43. Charlesworth, D., Charlesworth, B., 1979, A Model for the Evolution of Distyly, *The American Naturalist*, 111(4):467-98
 44. Charlesworth, D., Charlesworth, B., 1987, Inbreeding Depression and its Evolutionary Consequences, *Annual Review of Ecology and Systematics*, 18:237-68
 45. Chen, M., Loewus, F., 1977, *myo*-Inositol Metabolism in *Lilium longiflorum* Pollen, *Plant Physiology*, 59:653-7
 46. Chen, Y., Xu, Y., Luo, W., Li, W., Chen, N., Zhang, D., Chong, K., 2013, The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice, *Plant Physiology*, 163(4):1673-85
 47. Chen, F., Yang, Y., Li, B., Liu, Z., Khan, F., Zhang, T., Zhou, G., Tu, J., Shen, J., Yi, B., Fu, T., Dai, C., Ma, C., 2019, Functional Analysis of M-Locus Protein Kinase Revealed a Novel Regulatory Mechanism of Self-Incompatibility in *Brassica napus* L., *International Journal of Molecular Sciences*, 20(13):3303
 48. Chen, M., Fan, W., Hao, B., Zhang, W., Yan, M., Zhao, Y., Liang, Y., Liu, G., Lu, Y., Zhang, G., Zhao, Z., Hu, Y., Yang, S., 2020, *EbARC1*, an E3 Ubiquitin Ligase Gene in *Erigeron breviscapus*, Confers Self-Incompatibility in Transgenic *Arabidopsis thaliana*, *International Journal of Molecular Science*, 21(4):1458
 49. Chien, C., Bartel, P., Sternglanz, R., Fields, S., 1991, The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest, *Proceedings of the National Academy of Sciences of the United States of America*, 88(9578-82)

50. Ciechanover, A., Iwai, K., 2004, The Ubiquitin System: From Basic Mechanisms to the Patient Bed, *International Union of Biochemistry and Molecular Biology Life*, 656(4):193-201
51. Clontech Laboratories Incorporated, 2007, Matchmaker™ Library Construction & Screening Kits User Manual, Publication Number: PT3955-1 (PR742237)
52. Clusius, C., 1583, *Rariorum aliquot stirpium, per Pannoniam, Austriam, & vicinas quasdam provincias observatarum historia, quator libris expressa*. Antwerp, Belgium: Officina Chritophori Plantini
53. Cocker, J., Webster, M., Li, J., Wright, J., Kaithakottil, G., Swarbreck, D., Gilmartin, P., 2015, *Oakleaf*: an S locus-linked mutation of *Primula vulgaris* that affects leaf and flower development, *New Phytologist*, 208(1):149-61
54. Cocker, J., Wright, J., Li, J., Swarbreck, D., Dyer, S., Caccamo, M., Gilmartin, P., 2018, *Primula vulgaris* (primrose) genome assembly, annotation and gene expression, with comparative genomics on the heterostyly supergene, *Nature Scientific Reports*, 8:17942
55. Coen, E., Meyerowitz, M., 1991, The War of the Whorls: Genetic Interactions Controlling Flower Development, *Nature*, 353(6339):31-7
56. Cohen, J., 2016, *De novo* Sequencing and Comparative Transcriptomics of Floral Development of the Distylous Species *Lithospermum multiflorum*, *Frontiers in Plant Science*, 7:1934
57. Connelly, S., Manley, J., 1988, A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II, *Genes & Development*, 2:440-52
58. Cornish, E., Pettitt, J., Bonig, I., Clarke, A., 1987, Developmentally controlled expression of a gene associated with self-incompatibility in *Nicotiana glauca*, *Nature*, 326:99-102
59. Cronquist, A., 1968, *The Evolution and Classification of Flowering Plants*, Boston, USA: Houghton Mifflin Co.
60. Crosby, J., 1940, High proportions of homostyle plants in populations of *Primula vulgaris*, *Nature*, 145:672-3
61. Crosby, J., 1949, Selection of an unfavourable gene complex, *Evolution*, 3:212-30
62. Curtis, J., Curtis, C., 1985, Homostyle primroses re-visited. I. Variation in time and space, *Heredity*, 54:227-34
63. Curtis, R., Pankaj, Powers, S., Napier, J., Matthes, M., 2012, The *Arabidopsis* F-box/Kelch-Repeat Protein At2g44130 Is Upregulated in Giant Cells and Promotes Nematode Susceptibility, *PLoS ONE*, 7(1):e36433
64. Cyr, D., Langer, T., Douglas, M., 1994, DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70, *Trends in Biochemical Sciences*, 19(4):176-81
65. Dai, S., Chen, T., Chong, K., Xue, Y., Liu, S., Wang, T., 2007, Proteomics Identification of Differentially Expressed Proteins Associated with Pollen Germination and Tube Growth Reveals Characteristics of Germinated *Oryza sativa* Pollen, *Molecular & Cellular Proteomics*, 6(2):207-30
66. Darwin, C., 1862, On the two forms or dimorphic condition in the species of *Primula*, and on their remarkable sexual relations, *Journal of the Proceedings of the Linnean Society, Botany*, 6:77-96
67. Darwin, C., 1863, On the existence of two forms, and on their reciprocal sexual relation, in several species of the genus *Linum*, *Journal of the Proceedings of the Linnean Society (Botany)*, 7:69-83
68. Darwin, C., 1864, On the Sexual Relations of the Three Forms of *Lythrum salicaria*, *Journal of the Linnean Society of London (Botany)*, 8:169-96
69. Darwin, C., 1868, On the Specific Difference between *Primula veris*, Brit. Fl. (var. *officinalis*, of Linn.), *P. vulgaris*, Brit. Fl. (var. *acaulis*, Linn.) and *P. elatior*, Jacq.; and on the hybrid nature of the common Oxlip. With supplementary remarks on naturally-produced Hybrids in the genus *Verbascum*, *Journal of the Linnean Society of London*, 10:437-54
70. Darwin, C., 1876, *The effects of cross and self fertilisation in the vegetable kingdom*, London: John Murray
71. Darwin, C., 1877, *The Different Forms of Flowers on Plants of the Same Species*, John Murray, London
72. Darwin, F., 1887, *The life and letters of Charles Darwin, including an autobiographical chapter*, London, UK: John Murray
73. Davuluri, R., Suzuki, Y., Sugano, S., Zhang, M., 2000, CART classification of human 5'-UTR sequences, *Genome Research*, 10:1807-16.
74. de Framond, A., Barton, K., Chilton, M., 1983, Mini-Ti: a new vector strategy for plant genetic engineering, *Biotechnology*, 5:262-9
75. de Nettancourt, D., 1977, *Incompatibility in angiosperms*, Monographs on Theoretical and Applied Genetics 3, Springer-Verlag Berlin Heidelberg
76. Derbyshire, P., Ménard, D., Green, P., Sallbach, G., Buschmann, H., Lloyd, C., Pesquet, E., 2015, Proteomic Analysis of Microtubule Interacting Proteins over the Course of Xylem Tracheary Element Formation in *Arabidopsis*, *The Plant Cell*, 27:2709-26

77. Dickinson, H., 1995, Dry stigmas, water and self-incompatibility in *Brassica*, *Sexual Plant Reproduction*, 8:1-10
78. Dinneny, J., Weigel, D., Yanofsky, M., 2006, *NUBBIN* and *JAGGED* define stamen and carpel shape in *Arabidopsis*, *Development*, 133:1645-55
79. Dodds, P., Ferguson, C., Clarke, A., Newbikin, E., 1999, Pollen-expressed S-RNases are not involved in self-incompatibility in *Lycopersicon peruvianum*, *Sexual Plant Reproduction*, 12:76-87
80. Dowd, P., McCubbin, A., Wang, X., Verica, J., Tsukamoto, T., Ando, T., Kao, T., 2000, Use of *Petunia inflata* as a Model for the Study of Solanaceous Type Self-Incompatibility, *Annals of Botany*, 85(1):87-93
81. Du, J., Yan, L., Xiong, J., Ge, C., Iqbal, S., Qiao, Y., 2019, Identifying Genome-Wide Sequence Variations and Candidate Genes Implicated in Self-Incompatibility by Resequencing *Fragaria viridis*, *International Journal of Molecular Science*, 20(5):1039
82. Dulberger, R., 1964, Flower dimorphism and self-incompatibility in *Narcissus tazetta* L., *Evolution*, 18:361-3
83. East, E., 1940, The Distribution of Self-Sterility in the Flowering Plants, *Proceedings of the American Philosophical Society*, 82(4):449-518
84. Ebert, P., Anderson, M., Bernatzky, R., Altschuler, M., Clarke, A., 1989, Genetic polymorphism of self-incompatibility in flowering plants, *Cell*, 56(2):255-62
85. El-Maarouf-Bouteau, H., Sajjad, Y., Bazin, J., Langlade, N., Critescu, S., Balzergue, S., Baudouin, E., Bailly, C., 2014, Reactive oxygen species, abscisic acid and ethylene interact to regulate sunflower seed germination, *Plant, Cell & Environment*, 38(2):364-74
86. Ernst, A., 1928, Zur Vererbung der morphologischen Heterostylie merkmale, *Berichte der Deutschen Botanischen Gesellschaft*, 46:573-88
87. Ernst, A., 1933, Weitere untersuchungen zur Phänanalyse zum Fertilitätsproblem und zur Genetik heterostyler Primeln. II. *Primula hortensis*, *Archive der Julius Klaus Stiftung für Vererbungsforschung Sozialanthropologie und Rassenhygiene*, 11:1-280
88. Ernst, A., 1936, Heterostylie-Forschung Versuche zur genetischen analyse eines organisations und 'Anpassungs' merkmals, *Zeitschrift für Induktive Abstammungs und Vererbungslehre*, 71(1):156-230
89. Ernst, A., 1936, Erblchkeitsforschungen an calycanthemen Primeln, *Theoretical and Applied Genetics*, 8:313-24
90. Ernst, A., 1955, Self-fertility in monomorphic Primulas, *Genetica*, 27(1):391-448
91. Feldman, R., Correll, C., Kaplan, K., Deshaies, R., 1997, A Complex of Cdc4p, Skp1p, and Cdc53p/Cullin Catalyzes Ubiquitination of the Phosphorylated CDK Inhibitor Sic1p, *Cell*, 91(2): 221-30
92. Fields, S., Song, O., 1989, A novel genetic system to detect protein-protein interactions, *Nature*, 340(6230):245-6
93. Fischer, C., Kugler, A., Hoth, S., Dietrich, P., 2013, An IQ Domain Mediates the Interaction with Calmodulin in a Plant Cyclic Nucleotide-Gated Channel, *Plant & Cell Physiology*, 54(4):573-84
94. Franciosini, A., Lombardi, B., Lafrate, S., Pecce, V., Mele, G., Lupacchini, L., Rinaldi, G., Kondou, Y., Gusmaroli, Aki, S., Tsuge, T., Deng, X., Matsui, M., Vittorioso, P., Costantino, P., Serino, G., 2013, The *Arabidopsis* COP9 SIGNALOSOME INTERACTION F-BOX KELCH 1 Protein Forms an SCF Ubiquitin Ligase and Regulates Hypocotyl Elongation, *Molecular Plant*, 6(5):1616-29
95. Franklin-Tong, V., Thorlby, G., Lawrence, M., Franklin, C., 1992, Recognition, Signals, and Pollen Responses in the Incompatibility Reaction in *Papaver rhoeas*, *Angiosperm Pollen and Ovules*, Springer-Verlag
96. Franklin-Tong, V., Hacket, G., Hepler, P., 1997, Ratio-imaging of Ca²⁺ in the self-incompatibility response in pollen tubes of *Papaver rhoeas*, *The Plant Journal*, 12:1375-86
97. Frugier, F., Poirier, S., Satiat-Jeuenaître, Kondorosi, A., Crespi, M., 2000, A Krüppel-like zinc finger protein is involved in nitrogen-fixing root nodule organogenesis, *Genes & Development*, 14:475-82
98. Gagne, J., Downes, B., Shiu, S., Durski, A., Vierstra, R., 2002, *Proceedings of the National Academy of Sciences of the United States of America*, 99(17):11519-24
99. Gamas, P., Niebel, F., Lescure, N., Cullimore, J., 1996, Use of a Subtractive Hybridisation Approach to Identify New *Medicago trunculata* Genes Induced During Root Nodule Development, *Molecular Plant-Microbe Interactions*, 9(4):233-42
100. Ganders, F., 1979, The biology of heterostyly, *New Zealand Journal of Botany*, 17:607-35
101. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R., Bairoch, A., 2003, ExpASY: The proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Research*, 31(13):3784-8
102. Geiger, J., Jin, X., 2006, The structure and mechanism of myo-inositol-1-phosphate synthase, *Subcellular Biochemistry*, 39:157-80

103. Gilmartin, P., 2015, On the origins of observations of heterostyly in *Primula*, *New Phytologist*, 208:39-51
104. Gingras, A., Raught, B., Sonenberg, N., 1999, eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulations of Translation, *Annual Review of Biochemistry*, 68:913-63
105. Girin, T., Stephenson, P., Goldsack, C., Kempin, S., Perez, A., Pires, N., Sparrow, P., Wood, T., Yanofsky, M., Østergaard, L., 2010, Brassicaceae *INDEHISCENT* genes specify valve margin cell fate and repress replum formation, *The Plant Journal*, 63:329-38
106. Goldstein, G., Scheid, M., Hammerling, U., Boyse, E., Schlesinger, D., Niall, H., 1975, Isolation of a Polypeptide That Has Lymphocyte-Differentiating Properties and Is Probably Represented Universally in Living Cells, *Proceedings of the National Academy of Sciences of the United States of America*, 72(1):11-15
107. Guan, Y., Huang, X., Zhu, J., Gao, J., Zhang, H., Yang, Z., 2008, RUPTURED POLLEN GRAIN1, a member of the MtN3/saliva gene family, is a crucial for exine pattern formation and cell integrity of microspores in *Arabidopsis*, *Plant Physiology*, 147(852-63)
108. Gupta, V., Beggs, A., 2014, Kelch proteins: emerging roles in skeletal muscle development and diseases, *Skeletal Muscle*, 4:11
109. Haar, E., Musacchio, A., Harrison, S., Kirchhausen, T., 1998, Atomic Structure of Clathrin: A β Propeller Terminal Domain Joins an α Zigzag Linker, *Cell*, 95(4):563-73
110. Han, Y., Tang, A., Wan, H., Zhang, T., Cheng, T., Wang, J., Wang, W., Pan, H., Zhang, Q., 2018, An *APETALA2* Homolog, *RcAP2*, Regulated the Number of Rose Petals Derived From Stamens and Response to Temperature Fluctuations, *Frontiers in Plant Science*, 12(9):481
111. Hartmann-Petersen, R., Seeger, M., Gordon, C., 2003, Transferring substrates to the 26S proteasome, *Trends in Biochemical Sciences*, 28(1):26-31
112. Hassan, M., Zainal, Z., Ismail, I., 2015, Plant kelch containing F-box proteins: structure, evolution and functions, *RSC Advances*, 53:42808-14
113. Hayta, S., Smedley, M., Li, J., Harwood, W., Gilmartin, P., 2016, Plant Regeneration from Leaf-derived Callus Cultures of Primrose (*Primula vulgaris*), *HortScience*, 51(5):558-62
114. Hayta, S., Smedley, M., Li, J., Harwood, W., Gilmartin, P., 2018, *Agrobacterium*-mediated transformation systems of *Primula vulgaris*, *Plant Methods*, 14:93
115. He, L., Wu, W., Zinta, G., Yang, L., Wang, D., Liu, R., Zhang, H., Zheng, Z., Huang, H., Zhang, Q., Zhu, J., 2018, A naturally occurring epiallele associates with leaf senescence and local climate adaptation in *Arabidopsis* accessions, *Nature Communications*, 9:460
116. Hellens, R., Edwards, E., Leyland, N., Bean, S., Mullineaux, P., 2000, pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation, *Plant Molecular Biology*, 42:819-32
117. Hehl, A., Vassella, E., Braun, R., Roditi, I., 1994, A conserved stem-loop structure in the 3'-untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*, *Proceedings of the National Academy of Sciences of the United States of America*, 91: 370-4
118. Hepworth, S., Klenz, J., Haughn, G., 2006, UFO in the *Arabidopsis* inflorescence apex is required for floral-meristem identity and bract suppression, *Planta*, 223(4):769-78
119. Hicke, L., 2001, Protein regulation by monoubiquitin, *Nature Reviews: Molecular Cell Biology*, 2(3):195-201
120. Hildebrand, F., 1866, Über den trimorphismus der blüthen in der gattung Oxalis. *Monatsber König Preuss Akad Wiss Berl*, 52-374
121. Hoekama, A., Hirsch, P., Hooykas, P., Schilperoort, R., 1983, A binary plant vector strategy based on separation of cir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid, *Nature*, 303:179-80
122. Hohn, B., Hohn, T., 1982, Cauliflower Mosaic Virus: A Potential Vector for Plant Genetic Engineering, *Molecular Biology of Plant Tumors*, Academic Press, Section III, Chapter 22, 549-60
123. Hou, Y., Guo, X., Cyprys, P., Zhang, Y., Bleckmann, A., Cai, L., Huang, Q., Luo, Y., Gu, H., Dresselhaus, T., Dong, J., Qu, L., 2016, Maternal ENODLs Are Required for Pollen Tube Reception in *Arabidopsis*, *Current Biology*, 26(17):2343-50
124. Huang, S., Lee, H., Karunanandaa, B., Kao, T., 1994, Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen, *Plant Cell*, 6:1021-8
125. Huu, C., Kappel, C., Keller, B., Sicard, A., Takebayashi, Y., Breuninger, H., Nowak, M., Bäurle, I., Himmelbach, A., Burkart, M., Ebbing-Lohaus, T., Sakakibara, H., Altschmied, L., Conti, E., Lenhard, M., 2016, Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses, *eLife*, 5:e17956

126. Hsue, B., Masters, P., 1997, A Bulged Stem-Loop Structure in the 3' Untranslated Region of the Genome of the Coronavirus Mouse Hepatitis Virus Is Essential for Replication, *Journal of Virology*, 71(10):7576-8
127. Imaizumi, T., Schultz, T., Harmon, F., Ho, L., Kay, S., 2005, FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of *CONSTANS* in *Arabidopsis*, *Science*, 309(5732):293-7
128. Invitrogen, 2003, Gateway Technology User Guide, Document Number: 250522, Publication Number: MAN0000282
129. Invitrogen, 2001, NuPAGE® Novex high-performance pre-cast gels: Bis-Tris Gel Instruction Booklet, Publication Number: IM-8042E 010620
130. Invitrogen, 2013, One Shot® *ccdB* Survival™ 2 T1® Chemically Competent Cells, Publication Number: MAN0000761
131. Invitrogen, 2015, One Shot® TOP10 Chemically Competent *E. coli*, Publication Number: MAN0001491
132. Ioerger, T., Clark, A., Kao, T., 1990, Polymorphism at the self-incompatibility locus in Solanaceae predates speciation, *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9732-5
133. Ishimizu, T., Norioka, S., Kanai, M., Clarke, A., Sakyama, F., 1996, Location of cysteine and cystine residues in S-ribonuclease associated with gametophytic self-incompatibility, *European Journal of Biochemistry*, 242:627-35
134. Iwano, M., Ito, K., Fujii, S., Kakita, M., Asano-Shimosato, H., Igarashi, M., Kaothien-Nakayama, P., Entani, T., Kanatani, A., Takehisa, M., Tanaka, M., Komatsu, K., Shiba, H., Nagai, T., Miyawaki, A., Isogai, A., Takayama, S., 2015, Calcium signalling mediates self-incompatibility response in the Brassicaceae, *Nature Plants*, 1:15128
135. Jackson, J., Jones, G., Linskens, H., 1982, Phytic acid in pollen, *Phytochemistry*, 21:1255-8
136. Jackson, J., Kamboj, R., Linskens, H., 1983, Localization of phytic acid in the floral structure of *Petunia hybrida* and relation to the incompatibility genes, *Theoretical and Applied Genetics*, 64(3):259-62
137. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., Smale, S., 1994, DNA sequence requirements for transcriptional initiator activity in mammalian cells, *Molecular and Cellular Biology*, 14(1):16-27
138. Jefferson, R., Kavanagh, T., Bevan, M., 1987, GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants, *The Embo Journal*, 6(13):3901-7
139. Jia, Y., Gu, H., Wang, X., Chen, Q., Shi, S., Zhang, J., Ma, L., Zhang, H., Ma, H., 2012, Molecular cloning and characterization of an F-box family gene *CarF-box1* from chickpea (*Cicer arietinum* L.), *Molecular Biology Reports*, 39(3):2337-45
140. Jia, Y., Zhang, Q., Pan, H., Wang, S., Liu, Q., Sun, L., 2014, Callus induction and haploid plant regeneration from baby primrose (*Primula forbesii* Franch.) anther culture, *Scientia Horticulturae*, 176:273-81
141. Jiang, X., Gao, Y., Zhou, H., Chen, J., Qu, G., Zhang, S., 2014, Apoplastic calmodulin promotes self-incompatibility pollen tube growth by enhancing calcium influx and reactive oxygen species concentration in *Pyrus pyrifolia*, *Plant Cell Reports*, 33(2):255-63
142. Jimenez, A., Mansour, H., Keller, B., Conti, E., 2013, Low genetics diversity and high levels of inbreeding in the Sinai primrose (*Primula boveana*), a species on the brink of extinction, *Plant Systematics and Evolution*, 300(5):1199-208
143. Jordan, N., Franklin, F., Franklin-Tong, V., 2000, Evidence for DNA fragmentation triggered in the self-incompatibility response of *Papaver rhoeas*, *The Plant Journal*, 23:471-9
144. Juven-Gershon, T., Kadonaga, J., 2010, Regulation of gene expression via the core promoter and the basal transcriptional machinery, *Developmental Biology*, 15: 225-9
145. Kakeda, K., Jordan, N., Conner, A., Ride, P., Franklin-Tong, V., Franklin, F., 1998, Identification of residues in a hydrophilic loop of the *Papaver rhoeas* S protein that play a crucial role in the recognition of incompatible pollen, *Plant Cell*, 10:1723-31
146. Kao, T., McCubbin, A., 1996, How flowering plants discriminate between self and non-self pollen to prevent inbreeding, *Proceedings of the National Academy of Sciences of America*, 93:12059-65
147. Kappel, C., Huu, C., Lenhard, M., 2017, A short story gets longer: recent insights into the molecular basis of heterostyly, *Journal of Experimental Botany*, 68(21-22):5719-30
148. Karunanandaa, B., Huang, S., Kao, T., 1994, Carbohydrate moiety of the *Petunia inflata* S₃ protein is not required for self-incompatibility interactions between pollen and pistil, *Plant Cell*, 6:1933-40
149. Kawamura, Y., Uemura, M., 2003, Mass spectrometric approach for identifying putative plasma membrane proteins of *Arabidopsis* leaves associated with cold acclimation, *The Plant Journal*, 36(2):141-54
150. Kawarazaki, T., Kimura, S., Lizuka, A., Hanamata, S., Nibori, H., Michikawa, M., Imai, A., Abe, M., Kaya, H., Kuchitsu, K., 2013, A low temperature-inducible protein AtSRC2 enhances the ROS-producing

- activity of NADPH oxidase AtRbohF, *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research*, 1833(12):2775-80
151. Kawata, Y., Sakiyama, F., Tamaoki, H., 1988, Amino acid sequence of ribonuclease T2 from *Aspergillus oryzae*, *European Journal of Biochemistry*, 176(3):683-97
 152. Kenrick, J., Kaul, V., Williams, E., 1986, Self-incompatibility in *Acacia retinodes*: Site of pollen-tube arrest in the nucellus, *Planta*, 169(2):245-50
 153. Kent, O., 2016, Analysis of a candidate gene for the control of floral heteromorphy in *Primula vulgaris*, PhD Thesis, University of East Anglia, Norwich, UK
 154. Khosravi, D., Eric, C., Yang, C., Siu, M., Shore, J., 2004, High Level of α -Dioxygenase in Short Styles of Distylous *Turnera* Species, *International Journal of Plant Sciences*, 165(6):995-1006
 155. Kim, H., Chiang, T., Kieber, J., Schaller, G., 2013a, SCF^{KMD} controls cytokinin signalling by regulating the degradation of type-B response regulators, *Proceedings of the National Academy of Sciences of the United States of America*, 110(24):10028-33
 156. Kim, J., Dotson, B., Rey, C., Lindsey, J., Bleecker, A., Binder, B., Patterson, S., 2013b, New Clothes for the Jasmonic Acid Receptor *CO1*: Delayed Abscission, Meristem Arrest and Apical Dominance, *Public Library of Sciences One*, 8(4):e60505
 157. Kim, H., Kieber, J., Schaller, G., 2013c, The rice F-box protein KISS ME DEADLY 2 Functions as a negative regulator of cytokinin signalling, *Plant Signalling & Behaviour*, 8(12):e26434
 158. Kim, D., Langmead, B., Salzberg, S., 2015, HISAT: a fast spliced aligner with low memory requirements, *Nature Methods*, 12:357-60
 159. Kipreos, E., Pagano, M., 2000, The F-box protein family, *Genome Biology*, 1:reviews3002.1
 160. Kondo, K., Yamamoto, M., Matton, D., Sato, T., Hirai, M., Norioka, S., Hattori, T., Kowayama, Y., 2002, Cultivated tomato has defects in both *S-Rnase* and *HT* genes required for styler function of self-incompatibility, *The Plant Journal*, 29(5):627-36
 161. Kondo, K., McClure, B., 2008, New microsome-associated HT-family proteins from *Nicotiana* respond to pollination and define HT/NOD-24 protein family, *Molecular Plant*, 1(4):634-44
 162. Koornneef, M., Meinke, D., 2010, The development of *Arabidopsis* as a model plant, *The Plant Journal*, 61:909-21
 163. Koressaar, T., Remm, M., 2007, Enhancements and modifications of primer design program Primer3, *Bioinformatics*, 23(1):1289-91
 164. Kozak, M., 1986, Influences of mRNA secondary structure on initiation by eukaryotic ribosomes, *Proceedings of the National Academy of Science of the United States of America*, 83(9):2850-4
 165. Kozak, M., 1989, Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs, *Molecular and Cellular Biology*, 9(11):5134-42
 166. Kozak, M., 1991, An Analysis of Vertebrate mRNA Sequences: Intimations of Translational Control, *The Journal of Cell Biology*, 115(4):887-903
 167. Kozak, M., 1994, Features in the 5' non-coding sequences of rabbit alpha beta-globin mRNAs that affect translational efficiency, *Journal of Molecular Biology*, 235(1):95-110
 168. Kozak, M., 1999, Initiation of translation in prokaryotes and eukaryotes, *Gene*, 234:187-208
 169. Kubo, K., Entani, T., Takara, A., Wang, N., Fields, A., Hua, Z., Toyoda, M., Kawashima, S., Ando, T., Isogai, A., Kao, T., Takayama, S., 2010, Collaborative Non-Self Recognition System in *S*-RNase-Based Self-Incompatibility, *Science*, 330(6005):796-9
 170. Kubo, K., Tsukahara, M., Fujii, S., Murase, K., Wada, Y., Entani, T., Iwano, M., Takayama, S., 2016, Cullin1-P is an Essential Component of Non-Self Recognition System in Self-Incompatibility in *Petunia*, *Plant & Cell Physiology*, 57(11):2403-16
 171. Kurian, V., Richards, A., 1997, A new recombinant in the heteromorphy 'S' supergene in *Primula*, *Heredity*, 78:383-90
 172. Kutuzov, M., Andreeva, A., 2002, Protein Ser/Thr phosphatases with kelch-like repeat domains, *Cellular Signalling*, 14(9):745-50 atggaagt (11398-11719)
 173. Labonne, J., Goultiaeva, A., Shore, J., 2009, High-resolution mapping of the *S*-locus in *Turnera* leads to the discovery of three genes tightly associated with the *S*-alleles, *Molecular & General Genetics*, 281(6):673-85
 174. Labonne, J., Tamari, F., Shore, J., 2010, Characterization of X-ray-generated floral mutants carrying deletions at the *S*-locus of distylous *Turnera subulata*, *Heredity*, 105:235-43
 175. Labonne, J., Shore, J., 2011, Positional cloning of the *S* haplotype determining the floral and incompatibility phenotype of the long-styled morph of distylous *Turnera subulata*, *Molecular Genetics and Genomics*, 285(2):101-11

176. Lai, Z., Ma, W., Han, B., Liang, L., Zhang, Y., Hong, G., Xue, Y., 2002, An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum, *Plant Molecular Biology*, 50:29-41
177. Lambricht, D., Sondek, J., Boh, A., Skiba, N., Hamm, H., Sigler, P., 1996, The 2.0 Å crystal structure of a heterotrimeric G protein, *Nature*, 379:311-19
178. Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., Higgins, D., 2007, Clustal W and Clustal X version 2.0, *Bioinformatics*, 23(21):2947-8
179. Lee, H., Huang, S., Kao, T., 1994, *S* proteins control rejection of incompatible pollen in *Petunia inflata*, *Nature*, 367:560-3
180. Levin, J., Meyerowitz, E., 1995, *UFO*: An Arabidopsis Gene Involved in Both Floral Meristem and Floral Organ Development, *The Plant Cell*, 7(5):529-48
181. Lewin, B., 2001, *Genes VII*, Oxford University Press
182. Lewis, D., 1949, Incompatibility in flowering plants, *Biological reviews*, 24(4):472-96
183. Lewis, D., Jones, D., 1992, The genetics of heterostyly, *Evolution and function of heterostyly*, edited by Barrett, S., Berlin, Germany, published by Springer Verlag, pages 129-50
184. Li, X., Zhang, D., Hannink, M., Beamer, L., 2004, Crystal Structure of the Kelch Domain of Human Keap1, *Journal of Biological Chemistry*, 279:54750-8
185. Li, J., Webster, M., Furuya, M., Gilmartin, P., 2007, Identification and characterization of pin and thrum alleles of two genes that co-segregate with the *Primula S* locus, *The Plant Journal: for cell and molecular biology*, 51(1):18-31
186. Li, J., Webster, M., Dudas, B., Cook, H., Manfield, I., Davies, B., Gilmartin, P., 2008, The *S* locus-linked *Primula* homeotic mutant *sepaloid* shows characteristics of a B-function mutant but does not result from mutation in a B-function gene
187. Li, J., Dudas, B., Webster, M., Cook, H., Davies, B., Gilmartin, P., 2010, *Hose in Hose*, an *S* locus-linked mutant of *Primula vulgaris*, is caused by an unstable mutation at the *Globosa* locus, *Proceedings of the National Academy of Sciences of the United States of America*, 107(12):5664-8
188. Li, M., Tang, D., Wang, K., Lu, L., Yu, H., Gu, M., Yan, C., Cheng, Z., 2011a, Mutations in the F-box gene *LARGER PANICLE* improve the panicle architecture and enhance the grain yield in rice, *Plant Biotechnology Journal*, 9:1002-13
189. Li, J., Webster, M., Smith, M., Gilmartin, P., 2011b, Floral heteromorphy in *Primula vulgaris*: progress towards isolation and characterization of the *S* locus, *Annals of Botany*, 108:715-26
190. Li, J., Webster, M., Wright, J., Cocker, J., Smith, M., Badakshi, F., Heslop-Harrison, P., Gilmartin, P., 2015, Integration of genetic and physical maps of the *Primula vulgaris S* locus and localization by chromosome *in situ* hybridisation, *New Phytologist*, 208(1):137-48
191. Li, J., Cocker, J., Wright, J., Webster, M., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C., Gilmartin, P., 2016, Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*, *Nature Plants*, 2(12):16188
192. Liang, L., Lai, Z., Ma, W., Zhang, Y., Xue, Y., 2002, *AhSL28*, a senescence- and phosphate starvation-induced *S*-like RNase gene in *Antirrhinum*, *Biochimica et Biophysica Acta (BBA) – Gene Structure and Expression*, 1579(11):64-71
193. Linnaeus, C., 1792, *Botanicorum Principis, Philosophia Botanica*. London, UK: Andrew Murray
194. Linskens, H., 1997, Incompatibility in *Petunia*, *Proceedings of the Royal Society of B: Biological Sciences*, 188(1092):299-311
195. Liu, Y., Schiff M., Dinesh-Kumar, S., 2002, Virus-induced gene silencing in tomato, *The Plant Journal*, 31:777-86
196. Livak, K., Schmittgen, T., 2001, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, *Methods*, 25(4):402-8
197. Lloyd, D., Webb, C., 1992, The Evolution of Heterostyly, Chapter 6, in *Evolution and Function of Heterostyly*, Barrett, S., Monographs on Theoretical and Applied Genetics, volume 15, Springer, Berlin, Heidelberg,
198. Loewus, F., Murthy, P., 2000, *myo*-Inositol metabolism in plants, *Plant Science*, 150(1):1-19
199. Losada, A., Yokochi, T., Hirano, T., 2005, Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and *Xenopus* egg extracts, *Journal of Cell Science*, 118(10):2133-41
200. Love, S., Muir, T., Ramage, R., Shaw, K., Alexeev, D., Sawyer, L., Kelly, S., Price, N., Arnold, J., Mee, M., Mayer, J., 1997, Synthetic, structural and biological studies of the ubiquitin system: synthesis and crystal structure of an analogue containing unnatural amino acids, *Biochemistry Journal*, 323:727-34

201. Lu, Q., Tang, X., Tian, G., Wang, F., Liu, K., Nguyen, V., Kohalmi, S., Keller, W., Tsang, E., Harada, J., Rothstein, S., Cui, Y., 2010, Arabidopsis homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin, *The Plant Journal*, 61(2):259-70
202. Lu, W., Bian, X., Yang, W., Cheng, T., Wang, J., Zhang, Q., Pan, H., 2018, Transcriptomics Investigation into the Mechanisms of Self-Incompatibility between Pin and Thrum Morphs of *Primula maximowiczii*, *International Journal of Molecular Sciences*, 19(7):1840
203. Luu, D., Qin, X., Morse, D., Cappadocia, M., 2000, S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility, *Nature*, 407:649-51
204. Luukkonen, B., Tan, W., Schwartz, S., 1995, Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance, *Journal of Virology*, 69(7):4086-94
205. Macfarlane, S., 1999, Molecular biology of the tobnaviruses, *Journal of General Virology*, 80(11):2799-807
206. Makino, A., Mae, T., Ohira, K., 1984, Relation between Nitrogen and Ribulose-1,5-bisphosphate Carboxylase in Rice Leaves from Emergence through Senescence, *Plant and Cell Physiology*, 25(3):429-37
207. Manfield, I., Pavlov, V., Li, J., Cook, H., Hummel, F., Gilmartin, P., 2005, Molecular characterization of DNA sequences from the *Primula vulgaris* S-locus, *Journal of Experimental Botany*, 56(414):1177-88
208. Manzano, C., Abraham, Z., Lopez-Tórrejón, Pozo, J., 2008, Identification of ubiquitinated proteins in Arabidopsis, *Plant Molecular Biology*, 68(1-2):145-58
209. Mast, A., Kelso, S., Conti, E., 2006, Are any primroses (*Primula*) primitively monomorphic?, *New Phytologist*, 171:605-16
210. McClure, B., Mou, B., Canevascini, S., Bernatzky, R., 1999, A small asparagine-rich protein required for S-allele-specific pollen rejection in *Nicotiana*, *Proceedings of the National Academy of Sciences of the United States of America*, 96(23):13548-53
211. McClure, B., Franklin-Tong, V., 2006, Gametophytic self-incompatibility: understanding the cellular mechanisms involved in "self" pollen tube inhibition, *Planta*, 224(2):233-45
212. McCubbin, A., Chung, Y., Kao, T., 1997, A mutant S₃-RNase of *Petunia inflata* lacking RNase activity has an allele-specific dominant negative effect on self-incompatibility interactions, *Plant Cell*, 9:85-95
213. McCubbin, A., Lee, C., Hetrick, A., 2006, Identification of genes showing differential expression in developing flowers of *Primula vulgaris*, *Sexual Plant Reproduction*, 19(2):63-72
214. McCubbin, A., Kao, T., 2000, Molecular recognition and response in pistil interactions, *Annual Review of Cell and Developmental Biology*, 16(1):333-64
215. Meng, M., Geisler, M., Johansson, H., Harholt, J., Scheller, H., Mellerowicz, E., Kleczkowski, L., 2009, UDP-Glucose Pyrophosphorylase is not Rate Limiting, but is Essential in Arabidopsis, *Plant & Cell Physiology*, 50(5):998-1011
216. Mercer, A., Fleming, S., Ueda, N., 2005, F-Box-Like Domains are Present in Most Poxvirus Ankyrin Repeat proteins, *Virus Genes*, 31(2):127-33
217. Mignone, F., Gissi, C., Liuni, S., Pesole, G., 2002, Untranslated regions of mRNAs, *Genome Biology*, 3(3):reviews0004.1-0004.10
218. Montandon, C., Friso, G., Liao, J., Choi, J., Wijk, K., 2019, In Vivo Trapping of Proteins Interacting with the Chloroplast CLPC1 Chaperone: Potential Substrates and Adaptors, *Journal of Proteome Research*, 18(6):2585-600
219. Müller, G., Junnila, A., Butler, J., Kravchenko, V., Revay, E., Weiss, R., Schlein, Y., 2009, Efficacy of the botanical repellents geraniol, linalool, and citronella against mosquitos, *Journal of Vector Ecology*, 34(1):2-8
220. Murashige, T., Skoog, F., 1962, A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures, *Physiologia Plantarum*, 15:473-97
221. Murfett, J., Atherton, T., Mou, B., Gasser, C., McClure, B., 1994, S-RNase expressed in transgenic *Nicotiana* cause S-allele-specific pollen rejection, *Nature*, 367:563-6
222. Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T., Ishiguro, S., 2007, Improved Gateway Binary Vectors: High-Performance Vectors for Creation of Fusion Constructs in Transgenic Analysis of Plants, *Bioscience, Biotechnology and Biochemistry*, 71(8):2095-100
223. Nakamura, M., Tsunoda, T., Obokata, J., 2002, Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator, *The Plant Journal: for cell and molecular biology*, 29(1):1-10

224. Nasrallah, J., Kao, T., Goldberg, M., Nasrallah, M., 1985, A cDNA clone encoding an S-locus-specific glycol-protein from *Brassica oleraceae*, *Nature*, 318:263-7
225. Nekrutenko, A., Makova, K., Li, W., 2002, The K_A/K_S Ratio Test for Assessing the Protein-Coding Potential of Genomic Regions: An Empirical and Simulation Study, *Genome Research*, 12(1):198-202
226. Nelson, D., Lasswell, J., Rogg, L., Cohen, M., Bartel, B., 2000, *FKF1*, A Clock-Controlled Gene that Regulates the Transition to Flowering in *Arabidopsis*, 101(3):331-40
227. New England Biolabs, Phusion® High-Fidelity DNA Polymerase, Publication Number: M0530S
228. Nowak, M., Russo, G., Schlapbach, R., Huu, C., Lenhard, M., Conti, E., 2015, The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly, *Genome Biology*, 16:12
229. Ohnishi, O., Asano, N., 1999, Genetic diversity of *Fagopyrum homotropicum*, a wild species related to common buckwheat, *Genetic Resources and Crop Evolution*, 46(4):389-98
230. Oxley, D., Bacic, A., 1996, Disulphide bonding in a stylar self-incompatibility ribonuclease of *Nicotiana glauca*, *European Journal of Biochemistry*, 242:75-80
231. Pamela, V., Dowrick, J., 1955, Heterostyly and Homostyle in *Primula obconica*, *Heredity*, 10: 219-36
232. Pandey, 1959, Evolution of gametophytic and sporophytic systems of self-incompatibility in angiosperms, *Evolution*, 14(1), 98-115
233. Pandey, K., Troughton, J., 1974, Scanning electron microscopic observations of pollen grains and stigma in the self-incompatible heteromorphic species *Primula malacoides* FRANCH. and *Forsythia x intermedia* ZAB., and genetics of sporopollenin deposition, *Euphytica*, 23(2):337-44
234. Paton, A.J., Brummit, N., Govaerts, R., Harman, K., Hinchcliffe, S., Allkin, B. and Lughadha, E.N. (2008). Towards Target 1 of the Global Strategy for Plant Conservation: a working list of all known plant species—progress and prospects. *Taxon* 57(2): 1–10
235. Paquis, S., Mazeyrat-Gourbeyre, F., Fernandez, O., Crouzet, J., Clément, C., Baillieul, F., Dorey, S., 2011, Characterization of a F-box gene up-regulated by phytohormones and upon biotic and abiotic stresses in grapevine, *Molecular Biology Reports*, 38(5):3327-37
236. Peng, J., Schwartz, D., Elias, J., Thoreen, C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., Gygi, S., 2003, *Nature Biotechnology*, A proteomics approach to understanding protein ubiquitination, 21:921-6
237. Penin, A, Logacheva, M., 2011, Correlation between number and position of floral organs in *Arabidopsis*, *Annals of Botany*, 108(1):123-31
238. Perteua, M., Kim, D., Perteua, G., Leek, J., Salzberg, S., 2016, Transcript-level expression analysis of RNA-seq experiments with HISAT, Stringtie and Ballgown, *Nature Protocols*, 11:1650-67
239. Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., Liuni, S., 2001, Structural and functional features of eukaryotic mRNA untranslated regions, *Gene*, 276(1-2):73-81
240. Petegem, F., Chatelain, F., Minor, D., 2005, Insights into voltage-gated calcium channel regulation from the structure of the Cav1.2 IQ domain-Ca²⁺/calmodulin complex, *Nature Structural & Molecular Biology*, 12:1108-15
241. Pickart, C., 2001, Mechanisms Underlying Ubiquitination, *Annual Review of Biochemistry*, 70:503-33
242. Piper, J., Charlesworth, B., Charlesworth, D., 1984, A high rate of self-fertilisation and increased seed fertility of homostyle primroses, *Nature*, 310:50-1
243. Polverari, A., Molesini, B., Pezzotti, M., Buonauro, R., Marte, M., Delledonne, M., 2003, Nitric Oxide-Mediated Transcriptional Changes in *Arabidopsis thaliana*, *Molecular Plant-Microbe Interactions*, 16(12):1094-105
244. Porto, M., Pinheiro, N., Batista, V., Santos, R., Filho, P., Lima, L., 2014, Plant Promoters: An Approach of Structure and Function, *Molecular Biotechnology*: 56:38-49
245. Poulter, N., Vatovec, S., Franklin-Tong, V., 2008, Microtubules Are A Target for Self-Incompatibility Signaling in *Papaver* Pollen, *Cell Biology and Signal Transduction*, 146:1358-67
246. Poulter, N., Bosch, M., Franklin-Tong, V., 2011, Proteins implicated in mediating self-incompatibility-induced alterations to the actin cytoskeleton of *Papaver* pollen, *Annals of Botany*, 108(4):659-75
247. Prag, S., Adams, J., 2003, Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals, *BMC Bioinformatics*, 4:42
248. Prelich, G., 2012, Gene Overexpression: Uses, Mechanisms, and Interpretation, *Genetics*, 190(3):841-54
249. Promega, 2010, Wizard® Plus SV Minipreps DNA Purification System, Publication Number: TB225
250. Promega, 2018, GoTaq® Flexi DNA Polymerase, Publication Number: 9PIM829
251. Puerta, A., Ushikima, K., Koba, T., Sassa, H., 2009, Identification and functional analysis of pistil self-incompatibility factor *HT-B* of *Petunia*, *Journal of Experimental Botany*, 60(4):1309-18
252. Pulido, P., Llamas, E., Llorente, B., Ventura, S., Wright, L., Rodríguez-Concepción, 2016, Specific Hsp100 Chaperones Determine the Fate of the First Enzyme of the Plastidial Isoprenoid Pathway for Either

- Refolding or Degradation by the Stromal Clp Protease in Arabidopsis, Public Library of Science Genetics, 12(1):e1005824
253. QIAGEN, 2010, Critical Factors for Successful Real-Time PCR, Publication Number: 1063770
254. QIAGEN, 2018, QIAquick® Spin Handbook, Publication Number: HB-1196-004
255. QIAGEN, 2019, DNeasy® Plant Handbook, Publication Number: HB-1166-005
256. Qu, H., Zhang, Z., Wu, F., Wang, Y., 2016, The role of Ca²⁺ and Ca²⁺ channels in the gametophytic self-incompatibility of *Pyrus pyrifolia*, Cell Calcium, 60(5):299-308
257. Radzicka, A., Wolfenden, R., 1995, A proficient enzyme, Science, 267(5194):90-3
258. Ramegowda, V., Mysore, K., Senthil-Kumar, M., 2014, Virus-induced gene silencing is a versatile tool for unravelling the functional relevance of multiple abiotic-stress-responsive genes in crop plants, Frontiers in Plant Science, 5:323
259. Reichmann, J., Krizek, B., Meyerowitz, E., 1996, Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS, Proceedings of the National Academy of Sciences of the United States of America, 93:4793-8
260. Renault, L., Nassar, N., Vetter, I., Becker, J., Klebe, C., Roth, M., Wittinghofer, A., 1998, The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller, Nature, 392:97-101
261. Richards, J., 1993, Primula, Batsford Ltd, ISBN:0713470704
262. Richards, J., 1986, Plant Breeding Systems, London, UK: Allen & Unwin
263. Richards, J., 1997, Plant Breeding systems, 2nd edition, London, UK: Chapman & Hall
264. Rius, S., Casati, P., Iglesias, A., Gomez-Casati, D., 2008, Characterization of Arabidopsis Lines Deficient in GAPC-1, a Cytosolic NAD-Dependent Glyceraldehyde-3-Phosphate Dehydrogenase, Plant Physiology, 148:1655-67
265. Rockman, W., Wray, G., 2002, Abundant raw material for cis-regulatory evolution in humans, Molecular Biology and Evolution, 19(11):1991-2004
266. Rudd, J., Franklin, F., Lord, J., Franklin-Tong, C., 1996, Increased phosphorylation of a 26 kDa is induced by the self-incompatibility response, Plant Cell, 8:713-24
267. Rudd, J., Franklin, F., Lord, J., Frankling-Tong, V., 1997, Ca²⁺-independent phosphorylation of a 68 kDa protein is stimulated by the self-incompatibility response, The Plant Journal, 12:507-14
268. Running, M., Meyerowitz, E., 1996, Mutations in the *PERIANTHA* gene of *Arabidopsis* specifically alter floral organ number and initiation pattern, Development, 122:1261-9
269. Saiga, S., Moeller, B., Watanabe-Taneda, A., Abe, M., Weijers, D., Komeda, Y., 2012, Control of embryonic meristem initiation in Arabidopsis by PHD-finger protein complexes, Development, 139:1391-8
270. Sanchez-Serrano, J., Salinas, J., 2014, Arabidopsis Protocols, Methods in Molecular Biology (1062), Humana Press
271. Schopfer, C., Nasrallah, M., Nasrallah, J., 1999, The male determinant of self-incompatibility in *Brassica*, Science, 286:1697-700
272. Senthil-Kumar, M., Mysore, K., 2011, Virus-induced gene silencing can persist for more than 2 years and also be transmitted to progeny seedlings in *Nicotiana benthamiana* and tomato, Plant Biotechnology Journal, 9:797-806
273. Schindler, U., Menkens, A., Beckmann, H., Ecker, J., Chashmore, A., 1992, Heterodimerization between light-regulated and ubiquitously expressed Arabidopsis GBF bZIP proteins, 11:1261-73
274. Schindelman, G., Morkami, A., Jung, J., Baskin, T., Carpita, N., Derbyshire, P., McCann, M., Benfey, P., 2001, COBRA encodes a putative GPI-anchored protein, which is polarly localised and necessary for oriented cell expansion in *Arabidopsis*, Genes & Development, 15:1115-27
275. Schnurr, J., Storey, K., Jung, H., Somers, D., Gronwald, J., 2006, UDP-sugar pyrophosphorylase is essential for pollen development in Arabidopsis, Planta, 224(3):520-32
276. Schudoma, C., May, P., Nikiforova, V., Walther, D., 2010, Sequence-structure relationships in RNA loops: establishing the basis for loop homology modelling, Nucleic Acids Research, 38(3):970-80
277. Schultz, T., Kiyosue, T., Yanovsky, M., Wada, M., Kay, S., 2001, A Role for LKP2 in the Circadian Clock of Arabidopsis, The Plant Cell, 13:2659-70
278. Schumann, N., Navarros-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, 155:835-50
279. Schüpbach, T., Wieschaus, E., 1991, Female Sterile Mutations on the Second Chromosome of *Drosophila melanogaster*. II. Mutations Blocking Oogenesis or Altering Egg Morphology, Genetics, 129:1119-36

280. Serrano, I., Romero-Puertas, M., Sandalio, L., Olmedilla, A., 2015, The role of reactive oxygen species and nitric oxide in programmed cell death associated with self-incompatibility, *Journal of Experimental Botany*, 66(10):2869-76
281. Setyadjit, Irving, D., Joyce, D., Simons, D., 2006, Vase treatments containing gibberellic acid do not increase longevity of cut *Grevillea* 'Sylvia' inflorescences, *Australian Journal of Experimental Agriculture*, 46(11):1535-9
282. Shao, T., Qian, Q., Tang, D., Chen, J., Li, M., Cheng, Z., Luo, Q., 2012, A novel gene *IBF1* is required for the inhibition of brown pigment deposition in rice hull furrows, *Theoretical and Applied Genetics*, 125(2):381-90
283. Shao, J., Wang, H., Fang, S., Conti, E., Chen, Y., Zhu, H., 2019, Intraspecific variation of self-incompatibility in the distylous plant *Primula merrilliana*, *AoBP Plants*, 11(3):1-10, doi:10.1093/aobpla/plz030
284. Sievers, F., Wilm, A., Dineen, D., Gibson, T., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J., Higgins, D., 2011, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Molecular Systems Biology*, 7:539
285. Silva, N., Gorin, D., 2001, Mechanisms of self-incompatibility in flowering plants, *Cellular and Molecular Life Sciences CMLS*, 58:1988-2007
286. Silva, N., Stone, S., Christine, L., Sulaman, W., Nazarian, K., Burnett, L., Arnoldo, M., Rothstein, S., Goring, D., 2001, Expression of the S receptor kinase in self-compatible *Brassica napus* cv. Wester leads to the allele-specific rejection of self-incompatible *Brassica napus* pollen, *Molecular Genetics and Genomics*, 265(3):552-9
287. Skowyra, D., Craig, K., Tyers, M., Elledge, S., Harper, J., 1997, F-Box Proteins Are Receptors that Recruit Phosphorylated Substrates to the SCF Ubiquitin-Ligase Complex, *Cell*, 91(2):209-19
288. Smale, S., Kadonaga, J., 2003, The RNA polymerase II core promoter, *Annual Review of Biochemistry*, 72:449-79
289. Smalle, J., Vierstra, R., 2004, The Ubiquitin 26S Proteasome Proteolytic Pathway, *Annual Review of Plant Biology*, 55:555-90
290. Smedley, M., Harwood, W., 2015, Gateway®-Compatible Plant Transformation Vectors, *Methods in Molecular Biology*, 1223:3-16
291. Somssich, M., 2019, A short history of plant transformation, *PeerJ Preprints*, DOI: 10.7287/peerj.preprints.27556
292. Sorek, N., Sorek, H., Kijac, A., Szemenyei, H., Bauer, S., hematy, K., Wemmer, D., Somerville, C., 2014, The Arabidopsis COBRA protein facilitates cellulose crystallization at the plasma membrane, *Plant Cell*, 26(10):3491-20
293. Soulard, J., Boivin, N., Morse, D., Cappadocia, M., 2014, eEF1A Is an S-RNase Binding Factor in Self-Incompatible *Solanum chacoense*, *Public Library of Science One*, 9(2):e90206
294. Spitzer, N., 2008, Calcium: first messenger, *Nature Neuroscience*, 11:243-44
295. Sprengel, C., 1793, *Das entdeckte Geheimniss der Natur im Bau und in der Befruchtung der Blumen*, Berlin
296. Staiger, C., Franklin-Tong, V., 2003, The actin cytoskeleton is a target of the self-incompatibility response in *Papaver rhoeas*, *Journal of Experimental Botany*, 54(380):103-113
297. Stein, J., Howlett, D., Boyes, M., Nasrallah, M., Nasrallah, J., 1991, Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleraceae*, *Proceedings of the National Academy of Sciences of the United States of America*, 88(19):8816-20
298. Su, S., Dai, H., Wang, X., Wang, C., Zeng, W., Huang, J., Duan, Q., 2000, Ethylene negatively mediates self-incompatibility response in *Brassica rapa*, *Biochemical and Biophysical Research Communications*, doi:10.1016/j.bbrc.2020.02.128
299. Sudan, C., Prakash, S., Bhomkar, P., Jain, S., Bhalla-Sarin, N., 2006, Ubiquitous presence of β -glucuronidase (GUS) in plants and its regulation in some model plants, *Planta*, 224(4):852-64
300. Sultan, L., Milesina, D., Grewe, F., Rolle, K., Abudraham, S., Glodowicz, P., Niazi, A., Keren, I., Shevtsov, S., Klipcan, L., Barciszewski, J., Mower, J., Dietrich, A., Ostersetzer-Biran, O., 2016, The Reverse Transcriptase/RNA Maturase Protein MatR is Required for the Splicing of Various Group II Introns in Brassicaceae Mitochondria, *Plant Cell*, 28(11):2805-29
301. Sun, S., Footer, M., Matsudaira, P., 2017, Modification of Cys-837 identifies an actin-binding site in the beta-propeller protein scruin, *Molecular Biology of the Cell*, 8(3):387-545
302. Sun, Y., Zhou, X., Ma, H., 2007, Genome-wide Analysis of Kelch Repeat-containing F-box Family, *Journal of Integrative Plant Biology*, 49(6):940-52

303. Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., Suyama, A., Sugano, S., 2000, Statistical analysis of the 5'-untranslated region of human mRNA using "Oligo-Capped" cDNA libraries, *Genomics*, 64:286-97
304. Svoboda, P., Cara, A., 2006, Hairpin RNA: a secondary structure of primary importance, *Cellular and Molecular Life Sciences*, 901-18
305. Tang, W., Deng, Z., Osés-Prieto, J., Suzuki, N., Zhu, S., Zhang, X., Burlingame, A., Wang, Z., 2008, Proteomics Studies of Brassinosteroid Signal Transduction Using Prefractionation and Two-dimensional DIGE, *Molecular & Cellular Proteomics*, 7(4):728-38
306. Takahara, M., Magori, S., Soyano, T., Okamoto, S., Yoshida, C., Yano, K., Sato, S., Tabata, S., Yamaguchi, K., Shigenobu, S., Takeda, N., Suzaki, T., Kawaguchi, M., 2013, Too much love, a novel Kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume-*Rhizobium* symbiosis, *Plant Cell Physiology*, 54(4):433-47
307. Takahashi, T., Matsuhara, S., Abe, M., Komeda, Y., 2002, Disruption of a DNA Topoisomerase I Gene Affects Morphogenesis in Arabidopsis, *The Plant Cell*, 14:2085-93
308. Takahashi, N., Kuroda, H., Kuromori, T., Hirayama, T., Seki, M., Shinozaki, K., Shimada, H., Matsui, M., 2004, Expression and interaction analysis of Arabidopsis Skp1-related genes, *Plant Cell Physiology*, 45:83-91
309. Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., Hinata, K., 2000, The S receptor kinase determines self-incompatibility in *Brassica* stigma, *Nature*, 403:913-6
310. Takeshima, R., Nishio, T., Komatsu, S., Kurachi, N., Matsui, K., 2019, Identification of a gene encoding polygalacturonase expressed specifically in short styles in distylous common buckwheat (*Fagopyrum esculentum*), *Heredity*, doi: 10.1038/s41437-019-0227-x
311. Taylor, J., Whitelaw, C., 2001, Signals in abscission, *New Phytologist*, 151:323-39
312. The Arabidopsis Information Resource (TAIR), www.arabidopsis.org
313. Thermo Fisher Scientific, 2011, SYBR® Green PCR Master Mix and SYBR® Green RT-PCR Reagents Kit User Guide, Publication Number: 4310251
314. Thermo Fisher Scientific, 2014, SuperSignal® West Femto Maximum Sensitivity Substrate Instructions, Publication Number: B2160738
315. Thermo Fisher Scientific, 2016, Real-time PCR handbook, Publication Number: CO019861 0216
316. ThermoFisher Scientific, 2018, Qubit™ 4 Fluorometer, Invitrogen, Publication Number: MAN0017209
317. Thermo Fisher Scientific, 2018, High Capacity cDNA Reverse Transcription Kit User Guide, Publication Number: MAN0017977
318. Thiel, H., Hleibieh, K., Gilmer, D., Varrelmann, M., 2012, The P25 Pathogenicity Factor of *Beet necrotic yellow vein virus* Targets the Sugar Beet 26S Proteasome Involved in the Induction of a Hypersensitive Resistance Response via Interaction with an F-box Protein, *International Society for Molecular Plant-Microbe Interactions*, 25(8):1058-72
319. Tong, V., Hackett, G., Hepler, P., 1997, Ratio-imaging of Ca²⁺_i in the self-incompatibility response in pollen tubes of *Papaver rhoeas*, *The Plant Journal*, 12(6):1375-86
320. Tör, M., Mantell, S., Ainsworth, C., 1992, Endophytic bacteria expressing β-glucuronidase cause false positives in transformation of *Dioscorea* species, *Plant Cell Reports*, 11(9):452-6
321. Tsai, D., Lee, H., Post, L., Kreiling, K., Kao, T., 1992, Sequence of an S-protein of *Lycopersicon peruvianum* and comparison with other solanaceous S-proteins, *Sexual Plant Reproduction*, 5:256-63
322. Uetz, P., Dong, T., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S., Roupelieva, M., Rose, D., Fossum, E., Haas, J., 2006, Herpesviral Protein Networks and Their Interaction with the Human Proteome, *Science*, 311(5758):239-42
323. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., Rozen, S., Primer3 – new capabilities and interfaces, *Nucleic Acids Research*, 40(15):e115
324. Unver, T., Budak, H., 2009, Virus-Induced Gene Silencing, a Post Transcriptional Gene Silencing Method, *International Journal of Plant Genomics*, DOI: 10.1155/2009/198680
325. Ushijima, K., Nakano, R., Bando, M., Shigezane, Y., Ikeda, K., Namba, Y., Kume, S., Kitabata, T., Mori, H., Kubo, Y., 2012, Isolation of the floral morph-related genes in heterostylous flax (*Linum grandiflorum*): the genetic polymorphism and the transcriptional and post-transcriptional regulations of the S locus, *The Plant Journal: for cell and molecular biology*, 69(2):317-31
326. Vanoosthuysse, V., Tichtinsky, G., Dumas, C., Gaude, T., Cock, J., 2003, Interaction of Calmodulin, a Sorting Nexin and Kinase-Associated Protein Phosphatase with the *Brassica oleracea* S Locus Receptor Kinase, *Plant Physiology*, 133:919-29

327. Vierstra, R., 1993, Protein Degradation in Plants, *Annual Review of Plant Physiology and Molecular Biology*, 44:385-410
328. Vierstra, R., 1996, Proteolysis in plants: mechanisms and functions, *Post-Transcriptional Control of Gene Expression in Plants*, 275-302
329. Vierstra, R., 2003, The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins, *Trends in Plant Science*, 8(3):135-42
330. Vijay-Kumar, S., Bugg, C., Cook, W., 1987, Structure of ubiquitin refined at 1.8Å resolution, *Journal of Molecular Biology*, 194(3):531-44
331. Voges, D., Zwickl, P., Baumeister, W., 1999, The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis, *Annual Review of Biochemistry*, 68:1015-68
332. Voinnet, O., 2001, RNA silencing as a plant immune system against viruses, *Trends in Genetics*, 17(8):449-59
333. Vyetrogon, K., Tebbij, F., Olson, D., Ross, A., Matton, D., 2007, A comparative proteome and phosphoproteome analysis of differentially regulated proteins during fertilization in the self-incompatible species *Solanum chacoense* Bitt., *Proteomics*, 7:232-47
334. Wang, L., Peng, H., Ge, T., Liu, T., Hou, X., Li, Y., 2013, Identification of differentially accumulating pistil proteins associated with self-incompatibility of non-heading Chinese cabbage, *Plant Biology*, 16:49-57
335. Wang, L., Wang, C., Ge, T., Wang, J., Liu, T., Hou, X., Li, Y., 2014, Expression analysis of self-incompatibility-associated genes in non-heading Chinese cabbage, *Genetics and Molecular Research*, 13(3):5025-35
336. Wang, Y., Zhang, W., Song, L., Zou, J., Su, Z., Wu, W., 2008, Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in *Arabidopsis*, *Plant Physiology*, 148(3):1201-11
337. Webster, M., Gilmartin, P., 2003, A comparison of early floral ontogeny in wild-type and floral homeotic mutant phenotypes of *Primula*, *Planta*, 216(6):903-17
338. Webster, M., Gilmartin, P., 2006, Analysis of late stage flower development in *Primula vulgaris* reveals novel differences in cell morphology and temporal aspects of floral heteromorphy, *New Phytologist*, 171(3):591-603
339. Wedderburn, F., Richards, A., 1990, Variation in within-morph incompatibility inhibition sites in heteromorphy *Primula* L., *New Phytologist*, 116:149-62
340. Weitzel, A., Wyatt, S., Parsons-Wingter, P., 2016, Microarray data analysis of space grown *Arabidopsis* leaves for genes important in vascular patterning, *American Society for Gravitational and Space Research*, <https://ntrs.nasa.gov/search.jsp?R=20170002042>
341. Wittenbach, V., 1979, Ribulose Bisphosphate Carboxylase and Proteolytic Activity in Wheat Leaves from Anthesis through Senescence, *Plant Physiology*, 64:884-7
342. Wright, L., Phillips, M., 2014, Measuring the activity of 1-deoxy-D-xylulose 5-phosphate synthase, the first enzyme in the MEP pathway, in plant extracts, *Methods in Molecular Biology*, 1153:9-20
343. Xue, F., Cooley, L., 1993, *Kelch* encodes a component of intercellular bridges in *Drosophila* egg chambers, *Cell*, 72(5):681-93
344. Yamamoto, Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K., Abe, T., 2007, Identification of plant promoter constituents by analysis of local distribution of short sequences, *BMC Genomics*, 8:67
- Yan, F., Zhou, H., Yue, M., Yang, G., Li, H., Zhang, S., Zhao, P., 2019, Genome-Wide Identification and Transcriptional Expression Profiles of the *F-box* Gene Family in Common Walnut (*Juglans regia* L.), *Forests*, 10(3):275
345. Yan, Q., Xi, X., Sun, Z., Fang, Y., 2017, Depletion of *Arabidopsis* SC35 and SC35-like serine/arginine-rich proteins affects the transcription and splicing of a subset of genes, *Public Library of Science Genetics*, 13(3):e1006663
346. Yang, B., Sugio, A., White, F., 2006, Os8N3 is a host disease-susceptibility gene for bacterial blight of rice, *Proceedings of the National Academy of Sciences of the United States of America*, 103(10503-8)
347. Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., Ota, T., 2012, *S-LOCUS EARLY FLOWERING 3* Is Exclusively Present in the Genomes of Short-Styled Buckwheat Plants that Exhibit Heteromorphic Self-Incompatibility, *Public Library Of Science One*, 7(2):e31264
348. Yasui, Y., Hirakawa, H., Ueno, M., Matsui, K., Katsube-Tanaka, T., Yang, S., Aii, J., Sato, S., Mori, M., 2016, Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes, *DNA Research*, 23(3):215-24

349. Yu, S., Kim, H., Yun, D., Suh, M., Lee, B., 2019, Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1, *Plant Molecular Biology*, 99(1-2):135-48
350. Yuan, J., Reed, A., Chen, F., Stewart, C., 2006, Statistical analysis of real-time PCR data, *BMC Bioinformatics*, 7:85
351. Yuan, S., Barrett, S., Li, C., Li, X., Xie, K., Zhang, D., 2019, Genetics of distyly and homostyly in a self-compatible *Primula*, *Heredity*, 122:110-9
352. Zhang, Z., Li, J., Zhao, X., Wang, J., Wong, G., Yu, J., 2006, KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging, *Genomics, Proteomics & Bioinformatics*, 4(4):259-63
353. Zhang, X., Ghou, M., Liu, C., 2013, *Arabidopsis* Kelch Repeat F-Box Proteins Regulate Phenylpropanoid Biosynthesis via Controlling the Turnover of Phenylalanine Ammonia-Lyase, *The Plant Cell*, 25:4994-5010
354. Zhao, Z., Luo, Z., Yuan, S., Mei, L., Zhang, D., 2019, Global transcriptome and gene co-expression network analyses on the development of distyly in *Primula oreodoxa*, *Heredity*, doi: 10.1038/s41437-019-0250-y
355. Zhou, C., Yin, Y., Dam, P., Xu, Y., 2010, Identification of Novel Proteins Involved in Plant Cell-Wall Synthesis Based on Protein-Protein Interaction Data, *Journal of Proteome Research*, 9(10):5025-37
356. Zu, G., Ma, H., Nei, M., Kong, H., 2009, Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification, *Proceedings of the National Academy of Sciences of the United States of America*, 106(3):835-40