Population structure and the molecular genetics of petal spot pigmentation in *Gorteria diffusa*

Róisín Louise Fattorini

Christ's College

University of Cambridge

March 2021

This dissertation is submitted for the degree of Doctor of Philosophy.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit of 60,000 for the Biology Degree Committee.

Róisín Louise Fattorini

Dedication

This thesis is dedicated to the memory of Greg Mellers, Aleix Gorchs Rovira, and Hannah Elizabeth Taylor. Greg was an incredibly kind and competent mentor throughout my PhD, I was privileged to have his guidance and I really enjoyed working with him as a team. Aleix was a hugely positive presence in the Plant Sciences graduate community, and Hannah's unfailing encouragement during our school days and into adulthood was an incredible source of support.

Acknowledgements

I would like to thank my supervisor Beverley Glover for the great opportunity to complete this PhD project, and for all of the support and guidance she provided throughout. I thank Matthew Dorling for all of his help and in particular for plant care and his expertise on growing *G. diffusa*. To Edwige Moyroud for her fantastic mentorship, friendship, and support both with practical training and scientific quandaries. Boris Delahaie provided me with invaluable research skills, interesting *Gorteria* discussions, and the chance to be part of very well organised and thoroughly enjoyable fieldwork in South Africa. Without Eva Herrero Serrano's guidance and expertise none of the yeast work would have been possible, and I could not have asked for a better mentor in this regard. To Greg Mellers, who played a huge role in my development as a scientist, extremely skilled and full of enthusiasm for our research. I thank Chiara Airoldi and Qi Wang for their help, particularly with molecular troubleshooting and bioinformatics, respectively. I would like to thank Lize Joubert who played a vital role in my scientific training at the beginning of the PhD, and for her continued support and friendship throughout. To Alice Fairnie for her friendship and support, the scientific ponderings during lockdown walks, and her very useful feedback on a thesis chapter draft. Thea Kongsted provided valuable advice on specific techniques and Farah Khojayori for her ongoing work in continuing the project experiments. The other Glover lab postdocs and PhD students definitely enhanced my experience and I am very grateful to all of them for the mutual support and laughter. In particular I would like to thank Jordan Ferria, Erin Cullen, Gabriela Doria, Chris Davis, Roman Kellenberger, Emily Bailes, and Alfonso Timoneda Monfort.

I thank Lionel Hill, Hester Sheehan, Matt Davey, and Pallavi Singh for their help on different technical aspects, and Emma Jackson for help in taking care of the plants. To my second supervisor and tutor Julian Hibberd and Nick Gay for extremely valuable project and PhD advice. The interactions with our South African collaborators have been hugely beneficial. In particular I would like to thank Allan Ellis for his expertise and guidance in population genetics work, and Jurene Kemp for all of her help and support during field work. I am very grateful to my funding body NERC for enabling me to undertake the PhD project.

I would like to thank my wonderfully supportive family and friends. To my Biology teacher Clare Kelly who sparked my interest. Finally, thanks to my Great Uncle Thomas Igoe from whom I inherited my love of science.

Abstract

Petal spots are aggregations of pigmented cells in distinct regions of the petal and are known to function in pollinator attraction across multiple systems. They occur within many plant lineages, but petal spots of the South African daisy species *Gorteria diffusa* are unusually complex. These elaborate structures are richly pigmented, deeply textured, and include three distinct cell types. *G. diffusa* petal spots function in attracting male bee-fly pollinators, in one of only two known cases of sexual deception outside of the Orchidaceae. *G. diffusa* is comprised of geographically discrete floral morphotypes, defined by extreme intraspecific variation in capitulum phenotype. Between-morphotype variation in the position and complexity of petal spots is associated with differential pollinator behavioural responses. As such, this system has much potential for understanding the molecular development of an ecologically relevant trait within a powerful intraspecific comparative framework.

The sexually deceptive Spring morphotype was the focus for this project, which had two main aims: to investigate Spring population genetic structure and to characterise the genes regulating anthocyanin pigmentation within petal spots. For the latter, two additional morphotypes were studied to improve the robustness of conclusions based on comparisons between spotted and plain petal tissues.

The anthocyanin cyanidin 3-glucoside was found to pigment *G. diffusa* ray floret petals. In petal spots there was a high proportion of malonated anthocyanin, that was absent from other petal regions. In the first comprehensive characterisation of a *G. diffusa* petal spot developmental pathway, a small family of subgroup 6 R2R3 MYB transcription factors (GdMYB8 proteins) were identified as potential petal spot anthocyanin regulators. The genes encoding these proteins were found to be upregulated within petal spots and induced ectopic anthocyanin production when stably transformed into *Nicotiana tabacum*. Potential downstream targets of GdMYB8 proteins within the anthocyanin synthesis pathway were identified. The expression patterns of genes encoding these enzymes, and the ability of GdMYB8 proteins to bind to promoter regions of the anthocyanin synthesis genes (in yeast), imply that GdMYB8 proteins are likely to regulate *G. diffusa* petal spot anthocyanin production through co-regulation of several anthocyanin synthesis enzymes.

Extending our developmental approach into one which addresses the evolution of petal spot regulators across morphotypes, requires fundamental understanding of the genetic nature of morphotypes and in-depth characterisation of their floral phenotypes. Toward this aim, a genotyping by sequencing analysis of genetic structure within the Spring morphotype was conducted, along with floral phenotypic measurements of the individuals sequenced. Limited floral trait variation was detected within the Spring morphotype, but there was no grouping of phenotypes by locality. The genetic analysis indicated strong isolation by distance patterns, hypothesised to be due to limited seed dispersal. These findings suggest that limited dispersal may be a key component contributing toward the evolution of *G. diffusa* floral morphotypes, pending further investigation.

Ultimately, this research enhances our understanding of the genetics underlying *G. diffusa* petal spot development. It also demonstrates that isolation by distance is a major determinant of gene flow within a subset of *G. diffusa*, providing a first insight into the mechanisms that may facilitate the evolution of extreme intraspecific variation.

Contents

List of figures and tables

List of abbreviations

BAP Benzyl-Amino-Purine
bHLH basic-Helix-Loop-Helix
BLAST Basic Local Alignment Search
bp base pair
CaMV 35S Cauliflower Mosaic Virus 35S promoter
cDNA complementary deoxyribonucleic acid
CTAB Cetylmethylammonium Bromide
DNA Deoxyribonucleic acid
dNTP deoxyribonucleotide triphosphate
EDTA Ethylenediaminetetraacetic acid
EMSA Electromobility shift assay
EtBr Ethidium bromide
GA3 Gibberellic acid
GBS Genotyping-by-sequencing
gDNA genomic deoxyribonucleic acid
IAA Indole-3-acetic acid
IPTG Isopropyl β-D-1-thiogalactopyranoside
LB Luria-**B**etrani (medium)
LC-MS Liquid chromatography mass-spectrometry
mRNA messenger ribonucleic acid
MS Murashige-Skoog
MYB Transcription factors containing a domain first identified in myeloblastosis virus
NAA 1-Naphthaleneacetic acid
NADPH Nicotinamide adenine dinucleotide phosphate
ODx Optical density at wavelength X
PCA Principal component analysis
PCR Polymerase chain reaction
PVP Polyvinylpyrrolidone
qRT-PCR Real-time reverse transcription polymerase chain reaction
R2R3 Repeat 2 and repeat 3 of the MYB DNA-binding domain
REML Residual maximum likelihood
RACE Rapid amplification of cDNA ends
SDS Sodium dodecyl sulphate
SNPs Single nucleotide polymorphisms
T-DNA transfer DNA
TF Transcription factor
UTR Untranslated region
WD-40 A tryptophan-aspartic acid (W-D) terminating motif of roughly 40 amino acids
X-Gal 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Y1H Yeast one-hybrid
3-AT 3-amino-1, 2, 4-triazole

1.Introduction

1.1 Angiosperm diversification and plant-pollinator interactions

1.1.1 The evolution of flowering plants

Flowering plants originated 140-250 Mya and are the most speciose plant lineage, comprising 304,000 named species and up to 156,000 unnamed species (Magallón et al. 2015; Pimm and Raven 2017; Silvestro et al. 2015; Vamosi et al. 2018). The flower has been referred to as a key innovation in the evolution of complex organisms; unique to angiosperms, flowers are considered pivotal to the success of this lineage (Chanderbali et al. 2016; Soltis et al. 2019). Floral features such as a closed carpel that enables fruit to develop from the mature ovary wall, and the process of double fertilisation, may provide innate advantages. These advantages include enabling gametic competition, the provision of a nutrient source (endosperm) for the embryo, and fruit as a means of dispersing seeds (Soltis et al. 2019). The angiosperm clade is described as having a 'propensity to diversify that is evolutionarily labile' (Davies et al. 2004), due to repeated shifts in the rate of diversification and multiple radiations over angiosperm evolution (Davies et al. 2004; Soltis and Soltis, 2004). This complex pattern of diversification is likely to result from a combination of factors operating in a clade-specific manner (Marazzi and Sanderson 2010; Onstein et al. 2014; Sauquet and Magallón 2018; Soltis et al. 2019). Proposed drivers of diversification include whole genome duplications, extrinsic physical conditions, ecological interactions (including mutualisms with pollinating species), and the evolution of key traits (Hughes and Atchison 2015; Moore and Donoghue 2007; Van der Niet and Johnson 2012; O'Meara et al. 2016; Sargent 2004; Soltis et al. 2019; Weber and Agrawal 2014).

While we lack a holistic understanding of angiosperm lineage diversification, the evolution of key innovations and morphological transitions has been found to drive diversification at smaller taxonomic scales (Givnish 2010; Soulebeau et al. 2015; Vamosi et al. 2018). When relative increases in phenotypic and functional diversity coincide with, or precede, higher level diversification shifts this can indicate adaptive radiation (Givnish 2015). Adaptive radiation can accelerate species diversification when ecological roles diversify and corresponding adaptations occur in different species within a lineage (Givnish 1997 in Givnish 2015). Many examples of adaptive radiations have been found at the genus and family level, including in Hawaiian lobeliads (Givnish 2010, 2015; Soulebeau et al. 2015). The lobeliad genus *Cyanea* (76 species) has undergone parallel adaptive radiations on different islands resulting in ecological diversity and species richness. A hierarchical adaptive radiation appears to have occurred, in which habitat is the determinant at the genus level, and subsequently elevation and flower-tube length drove adaptive radiations within *Cyanea* (Givnish et al. 2009). Understanding processes that drive adaptive radiations and the contribution of floral evolution to these events requires knowledge of the evolutionary and developmental mechanisms that determine differences in phenotypes, as well as the ecological consequences of these trait changes (Shan et al. 2019).

1.1.2 Pollinators and floral diversification

Floral variation is often attributed to pollinator mediated selection, which is considered an important driver of flowering plant evolution (Kay and Sargent 2009; Smith et al. 2018). Plants evolve to enhance reproductive success by attracting pollinators with floral displays and, as such, pollinators exert selection on floral traits including scent, colour, and flower shape. The association between flowering

plants and pollinators may have been co-opted from pre-existing gymnosperm-pollinator mutualisms (Labandeira 2010; Labandeira et al. 2007; Soltis et al. 2019). Animal pollination occurs in approximately 78-94% of flowering plants (Ollerton 2011). Bees are major pollinators of many flowering plant species, and the origin of the crown bees 123 Mya is thought to coincide with diversification of the eudicots, which contain 75% of angiosperm species (Cardinal and Danforth 2013; Hu et al. 2008). Specialisation between flowering plants and pollinators has been cited as a mechanism driving divergence (Armbruster et al. 2014; Armbruster and Muchhala 2009).

Spatial variation in pollinator assembly can cause divergent selection on plants in different geographical locations. Ongoing pollinator evolution results in shifting floral phenotypic optima, potentially leading to assortative mating. The consequent reduction in pollen transfer between plants with different phenotypes can promote floral isolation (Grant and Grant 1965; Grant 1949; Kay and Sargent 2009; Schemske 2009). The pollinator-shift model (Grant and Grant 1965; Stebbins 1970) explains that transitions between communities of pollinators on a macroevolutionary scale result in alterations to multiple floral traits (Smith and Kriebel 2018). Studies largely conducted within species and between species pairs have found strong relationships between the evolution of floral traits and pollination systems (e.g. Pérez-Barrales et al. 2007; Streisfeld and Kohn 2007). However, macroevolutionary analyses are required to determine how shifts between major pollinator groups contribute toward speciation events (Van der Niet and Johnson 2012; Van der Niet et al. 2014). One such study empirically investigated the relationship between pollination system and flower shape in *Iochroma* and findings were consistent with the pollinator-shift model. *Iochroma* contains species pollinated by hummingbirds, or both hummingbirds and insects, and insect pollination is also predominant in closely related genera (Taura and Laroca 2004; Verçoza et al. 2012 in Smith and Kriebel 2018). Narrow, tubular flowers primarily attract hummingbirds and are the ancestral state, while campanulate or open bowl-shaped flowers tend to attract insect pollinators. Multiple shifts occurred from the ancestral tubular flower state to the open forms that correlated with pollination system changes (Smith and Kriebel 2018). Floral traits such as nectar spurs, corolla tubes, and bilateral symmetry can facilitate specialisation by accommodating only specific pollinator morphologies. These innovations can trigger co-evolutionary arms races in which adaptations and counter-adaptations, of both plant and pollinator, lead to highly specialised interactions that restrict gene flow and contribute toward lineage divergence (Woźniak and Sicard 2018). It is important to note that radiations have also occurred in multiple florally diverse groups while the specialised pollination ecology has been retained (De Luca and Vallejo-Marin 2013; Davis et al. 2014). It is often synergy between floral isolation and other mechanisms, such as local adaptation and postzygotic isolation, that provides strong barriers to gene flow, thus driving speciation in angiosperm lineages (Kay and Sargent 2009).

1.1.3 Floral signalling

Plant reproductive success is reliant on the transfer of pollen to and from conspecific flowers. In reciprocally beneficial interactions, pollinators receive floral rewards during pollen collection and deposition that are advertised through floral displays. Nectar and pollen are floral rewards with nutritional value, and pollinators can also gain nesting materials, heat sources, and sites for sleeping, brooding, and mating (Balamurali et al. 2015). Effective communication is integral to the success of plant-pollinator mutualisms, with floral signals optimised to provide stimuli that inform the receiver and ensure detection of flowers within a noisy environment (Endler 1992). These stimuli can be visual, thermal, tactile, and olfactory. The context of olfactory stimuli can alter the information that is

transmitted. Stimulus presentation can differ depending on its concentration and timing, while receiver condition (e.g. gender and experience) can alter signal interpretation (Raguso 2008). Methyl salicylate is a plant volatile that exemplifies these context-dependent effects, acting as an attractant to orchid bees (Eltz et al. 2005), reducing honeybee visitation (Henning et al. 1992), and stimulating hunting behaviours in carnivorous mites that predate plant herbivores (De Boer and Dicke 2004). Complexity is also evident in signal properties, with visual signals varying in factors such as size, pattern, iridescence, colour, and symmetry (Leonard et al. 2011). The functioning of visual and olfactory stimuli, and the synergism between them, has been demonstrated through pollinator choice experiments (Klahre et al. 2011); for example, the hawkmoth *Manduca sexta* feeds on the nectar of *Datura wrightii* but requires both olfactory and visual cues to induce this feeding response (Raguso and Willis 2005). Floral signals from different modalities can also act redundantly - when pollinators do not exploit all sensory cues. Floral signal complementation occurs when, for example, one modality serves as a long distance attractant and the second enables precise location of the floral reward (Raguso 2004). The sensory capabilities of the pollinator determine whether floral traits can be detected and pollinator responses to these signals are derived from learning abilities, innate preferences, and pre-existing biases (Balamurali et al. 2015; Kelber et al. 2003; Peitsch et al. 1992). Some insect pollinators prefer floral volatiles that are also produced by insects, providing an example of potential exploitation by the plant of these pre-existing biases (Raguso 2008; Vlasáková et al. 2008). The relationship between floral signal and receiver interpretation has resulted in some specific floral attributes evolving to attract particular pollinator groups. However, this relationship is also influenced by the compatibility of pollinator physiology with particular floral forms (Shan et al. 2019).

1.1.4 Pollinator-attracting floral traits

Floral organs develop traits that are often crucial for pollinator attraction and defence against pathogens and florivores, including olfactory, visual, and gustatory features (Shan et al. 2019). A typical eudicot flower has four types of floral organ arranged in whorls: sepals, petals, stamens and carpels, respectively from the periphery to the centre of the receptacle. Alterations to these organs result in diverse floral phenotypes, for example, an increase or reduction in size of petals or stamens in particular positions can produce a bilaterally symmetrical (zygomorphic) flower. Evolutionary shifts to zygomorphy are traditionally associated with specialised pollination (Endress 2012; Fenster et al. 2004; Shan et al. 2019), supported by a recent meta-analysis of floral visitation networks showing that species with floral zygomorphy tended to have fewer floral visitors than species with actinomorphic (radially symmetric) flowers (Yoder et al. 2020). Changes in symmetry are necessary for the development of within-whorl heterogeneity, regarding floral organ shape and size. Altering floral organ dimensions, within or between floral whorls, can impact the overall flower shape and these differences in shape can be detected by bat and insect pollinators (Gómez et al. 2006; Muchhala and Serrano 2015; Yoshioka et al. 2007). Differences in flower size are also recognised by hummingbird and insect pollinators, with bumblebees exhibiting slower foraging between flowers of a smaller size (Brody 1992; Kaczorowski et al. 2012; Spaethe et al. 2001). The deformation of floral tissue caused by regional alterations to cell proliferation and cell expansion can produce highly specialised structures (Shan et al. 2019). Fusion (i.e. a lack of separation) of floral organs also provides new structural variation and, in *Campsis grandiflora* flowers, pollen removal is enhanced by fusion of the anthers (Ren and Tang 2010; Specht and Howarth 2015). Complex floral shapes can be advantageous to pollinators with compatible morphologies but prevent access to floral rewards for other species. Long nectar spurs, for example, are typically associated with pollination by moths with a proboscis long

enough to reach nectar at the base of the spur (Darwin 1862). In *Impatiens burtonii* the nectar spur acts as a resource for short-proboscid and long-proboscid pollinators that occupy different temporal and spatial niches; long-proboscid pollinator visitation increases as nectar levels decrease over the course of the day, and vice versa (Vlašánková et al. 2017).

During maturation of the floral organs, petal epidermal cells change shape influencing texture and colour, volatiles form floral scent profiles, and flower colouration develops through pigmentation and structural colour (Moyroud and Glover 2017). Flavonoids, including the pigment anthocyanin, may protect the flower by acting as feeding deterrents to herbivores and contributing toward plant stress tolerance by providing a shield against UV damage (Ferrer et al. 2008; Narbona et al. 2018; Tripp et al. 2018; War et al. 2012). It is well established that pollinator selection is affected by changes in flower colour in groups including moths, bees, and hummingbirds (Davies et al. 2012; Hoballah et al. 2005; Papiorek et al. 2016; Sheehan et al. 2012). Hummingbirds are major pollinators of the red-flowered *Mimulus cardinalis*, whereas *Mimulus lewisii* is pink-flowered and mainly bee-pollinated (Bradshaw Jr et al. 1995). Crossing these species to form hybrids, and then near-isogenic lines (NIL), produced *M. cardinalis* plants with dark pink flowers and *M. lewisii* plants with orange flowers (Bradshaw Jr and Schemske 2003). Bees preferred pink *M. cardinalis* flowers over red-flowered wild types and hummingbirds preferred orange *M. lewisii* flowers over the pink-flowered wild type (Bradshaw Jr et al. 1995; Schemske and Bradshaw 1999). This experiment demonstrated that floral colour influences pollinator preferences in two distantly related pollinator species. Patterns of pigmentation can also be important for pollinator attraction including stripes, bicolour, and spots (Eckhart et al. 2006; Gaskett 2011; Leonard et al. 2011; Moeller 2005; Shang et al. 2011). Petal spots are defined as discrete aggregations of pigmented cells that contrast with the background colouration of the flower. Petal spots may function in pollinator attraction in several ways including increasing floral temperature through heat absorption (Dyer et al. 2006), acting as tactile or visual nectar guides (Leonard and Papaj, 2011), making flowers more conspicuous (de Jager et al. 2017), or by triggering mating or aggregation behaviours (Ellis and Johnson 2010; Johnson and Midgley 1997). Dark petal spots have evolved independently multiple times and occur in many families including Orchidaceae, Fabaceae, Liliaceae, and Asteraceae (Martins et al. 2013). Differences in colouration between floral organs also produces patterning. *Commelina communis* L. have yellow anthers that stand out against blue petals, the anthers promote pollinator landing and appropriate orientation of the pollinator for pollination (Ushimaru et al. 2007). Among other signals, flower colouration and pigment patterning can be very important for pollinator attraction in a wide range of species.

1.2 Flower development

1.2.1 Plant molecular evolution

Plant developmental processes are regulated by a core subset of proteins and require intricate coordination of complex genetic interactions. Alterations to these pathways can significantly disrupt plant development, proving detrimental to the organism. On rare occasions, these alterations instead lead to the acquisition of new processes or patterns that result in novel functional phenotypes (Jiggins et al. 2016). Existing adaptations and historical contingency constrain evolution. Likened to a 'tinkerer', evolution opportunistically modifies existing systems, circumventing developmental constraints 'by using differently the same structural information' (Jacob 1977). The regulation of gene expression is an important control mechanism that dictates developmental processes, and enables

plants to respond to environmental stresses (Feller et al. 2011; Yang et al. 2012). As such, it is widely acknowledged that genetic diversification in regulatory regions of genes may be key to developmental evolution (Carroll 2008; Hoekstra and Coyne 2007; Prud'homme et al. 2007). Rapid morphological diversification has been associated with increased rates of regulatory gene evolution in plants (Barrier et al. 2001; Lang et al. 2010). Specifically, the formation of novel regulatory interactions through *cis*-regulatory evolution has been proposed as a predominant factor in phenotypic evolution (Wittkopp and Kalay 2012; Wray 2007). The *cis*-regulatory region of a gene is comprised of all DNA elements that regulate the expression of the gene-coding region directly, without encoding intermediary factors (Stern and Orgogozo 2008). C*is*-regulatory elements are regulatory DNA sequences that contain transcription factor binding sites (Wittkopp and Kalay 2012). The highly modular organisation of many *cis*-regulatory elements enables separation and modification of developmental components independent from other developmental processes (Ambrose and Ferrándiz 2018; Bolker 2000). Mutations within these regulatory sequences may, therefore, be less likely to generate pleiotropic effects detrimental to development, in comparison to mutations in coding regions (Carroll 2008; Rebeiz et al. 2015; Stern 2000). However, redundancy and modularity of gene regulatory networks may enable them to act as robust buffers also counteracting the potential negative pleiotropic effects of a coding mutation in a developmental gene (Garfield et al. 2013; Hoekstra and Coyne 2007; Prud'homme et al. 2007). From our understanding of the network governing *Arabidopsis thaliana* flower development it is clear that common transcription factors are essential process integrators in a 'fantastically intricate web of crosstalk, feedback, and redundancy' (Posé et al. 2012).

1.2.2 The role of transcription factors

Transcription factor proteins bind to sequence-specific regions of DNA and regulate the initiation and rate of target gene transcription; they can both activate and repress transcription, sometimes through interactions with other proteins (Feller et al. 2011; Lehti-Shiu et al. 2017; Yang et al. 2012). A significant proportion of protein-encoding genes function in the regulation of gene expression; in *Arabidopsis thaliana* transcription factors account for 6% of these genes, while in *Caenorhabditis elegans* they account for only 3.6% (Riechmann et al. 2000). On average, 1500-2000 transcription factor genes are found within sequenced plant genomes (Feller et al. 2011; De Mendoza et al. 2013; Mitsuda and Ohme-Takagi 2009; Riechmann et al. 2000; Yamasaki et al. 2013). The specific sites that transcription factors bind to occur in regulatory regions (often promoters) of target genes. They are involved in several mechanisms coordinating the regulation of gene expression, and RNA polymerase requires both general and specific transcription factors to bind DNA and synthesise RNA during transcription (Feller et al. 2011; Stracke et al. 2001). To regulate gene expression transcription factors can integrate environmental signals and internal signals, providing a complex control mechanism that both modulates developmental processes and responds to external stresses (Lehti-Shiu et al. 2017; Yang et al. 2012).

Transcription factors have a modular structure and have been classified into families based on similarities in the DNA-binding domain module (Stracke et al. 2001). A lot is known about the evolution of certain transcription factor families including APETALA2 /Ethylene Response Factor (AP2 /ERF)(Kim et al. 2006; Mizoi et al. 2012), MADS-box (Gramzow et al. 2010; Shan et al. 2009), and MYB families (Du et al. 2015; Feller et al. 2011; Yuan et al. 2014). Transcription factors are integral to plant responses to environmental challenges, regulating stress-responsive transcription (e.g. Heat shock factor (Ahn et al. 2001) and Zinc finger (Deng et al. 2014) proteins) (Nakai et al. 2013; Nakashima et al. 2012; Singh et al. 2002) and other physiological processes dependent on environmental cues, such as fruit

maturation (Bastías et al. 2014), circadian rhythms (Gendron et al. 2012), and flowering (Corbesier et al. 2007). Through regulation of metabolic enzymes, transcription factors have a key role in controlling secondary metabolite accumulation, impacting plant cellular metabolism (Lehti-Shiu et al. 2017; Yang et al. 2012). Large transcription factor families have been found to play key roles in the evolution of developmental processes, for example, MADS-box genes (Ambrose and Ferrándiz 2018). The copy number of transcription factors from individual transcription factor families is highly variable across different species. Transcription factor families that have expanded in specific plant lineages may regulate clade-specific functions (Dias et al. 2003; Shiu et al. 2005).

Regulatory changes in gene expression can evolve in several ways including de novo through mutations in *cis*-regulatory elements or via mutations in coding regions of transcription factors. The latter produces heritable regulatory changes by causing downstream alterations that impact gene expression spatially and/or temporally. Acquisition of a new transcription factor binding site can alter the expression domain of a gene by repressing existing expression or activating new expression at a particular time, location, or in certain conditions. Similarly, changes in gene expression can occur from loss of transcription factor binding sites (Carroll 2008; Wittkopp and Kalay 2012). Some transcription factor binding events do not appear to influence gene expression, they are perhaps remnants of past functions and some may require specific environmental conditions or genetic backgrounds in order to regulate gene expression (Li et al. 2008; Schmidt et al. 2010). Some genetic changes can also alter *cis*-regulatory and coding regions of one or multiple genes, for example, gene loss, gene rearrangement, or gene duplication (Stern and Orgogozo 2008). Gene duplication and polyploidy are commonplace within the plant lineage and are considered an important source of developmental variation. In model plant systems, paralogues functioning in transcriptional regulation are preferentially retained, suggesting that these processes may be important contributors to the evolution of novel form and function (Blanc and Wolfe 2004; Jiao and Paterson 2014; Rensing 2014). Following gene duplication, novel gene functions can result from shuffling or recombination of protein domains (Kersting et al. 2012). Expansion of transcription factor families tends to occur after genome duplication, which provides a mechanism to increase the complexity and number of gene regulatory networks and greater opportunity for network subfunctionalisation and neofunctionalisation. In plants there is a relatively higher rate of transcription factor duplicates retained relative to other lineages (Blanc and Wolfe 2004; Jiang et al. 2013; Shiu et al. 2005).

1.2.3 The genetics of flower development

Extensive floral variation has evolved within the constraints of the genetic network underpinning flower development. All floral organs undergo key developmental processes of initiation, identity determination, morphogenesis, and maturation. Despite sharing basic developmental mechanisms, each type of floral organ usually has a distinct trajectory during development. The culmination of differences in these processes produces variability within and between floral organs that can manifest as striking intraspecific diversity when considering the entire floral phenotype (Irish 2008; Moyroud and Glover 2017; Shan et al. 2019; Walcher-Chevillet and Kramer 2016). A combination of endogenous and environmental factors trigger floral development, when the shoot apical meristem transitions into an inflorescence meristem that eventually leads to production of a meristem for each individual flower (Krizek and Fletcher 2005). Floral organ primordia develop from a small number of founder cells at floral meristem peripheries and form in a centripetal sequence through cell proliferation (Chandler 2011; Chandler et al. 2011). The phytohormones auxin and cytokinin are important in specifying the precise location of floral organ initiation, which occurs at auxin maxima reinforced by a cytokinin

gradient (Besnard et al. 2014). In *Arabidopsis*, the transcription factor LEAFY (LFY) is a master regulator of the whole floral network, activating genes that produce floral meristem and floral organ primordia (Benlloch et al. 2007; Blázquez et al. 2006; Irish 2010; Liu et al. 2009; Moyroud et al. 2009; Wagner 2009). APETALA 1 (AP1) is a transcription factor that is also involved in activating these processes, directly targeted by LFY and also acting in parallel with LFY (Winter et al. 2015). In other angiosperm species both genes have conserved regulatory functions in certain aspects of flower development but not all (Ahearn et al. 2001; Berbel et al. 2001; Huijser et al. 1992; Kato et al. 2005; Molinero-Rosales et al. 1999; Moyroud et al. 2009; Rottmann et al. 2000; Souer et al. 1998, 2008; Taylor et al. 2002; Vrebalov et al. 2002). The *Antirrhinum* AP1 ortholog, for example, does not have any conspicuous role in the development of individual floral organs while in *Arabidopsis* AP1 is involved in sepal and petal development (Bowman et al. 1993; Irish and Sussex 1990, Huijser et al. 1992, Irish 2009).

Floral-meristem identity genes, including LFY and AP1, regulate the expression of floral homeotic genes that control floral organ identity. Specific combinations of these homeotic proteins are hypothesised to form tetrameric regulatory complexes in the organ primordia of floral meristem (Hugouvieux et al. 2019). These complexes are thought to function as transcription factors by binding to the DNA of target genes initiating and maintaining specific floral organ identities (Theißen 2001; Theißen and Saedler 1999, 2001). Mutants were identified that had certain floral organ identities replaced with another type of floral organ, termed homeosis. In *A. thaliana*, these floral homeotic mutants were categorised into A, B, and C classes (Bowman et al. 1991, Coen and Meyerowitz, 1991). In *Arabidopsis*, for example, AP1 is an A function protein and AGAMOUS is a C function protein. The (A)B(C) model provides a framework explaining how these different transcription factors specify organ identity and correct organ positioning. This model was extended through the addition of a D class that functions in ovule identity specification (Angenet and Colombo 1996) and, subsequently, *SEPALLATA*-like genes were found to form an E class of floral organ identity genes (Pelaz et al. 2000; Ditta et al. 2004; Theißen 2001). Typically, perianth organs develop from the oldest primordia, with sepals (the calyx) developing due to the activity of A and E class proteins and petals (the corolla) arising as a result of A and E proteins alongside B proteins. Stamens (the androecium) arise from the next primordia through a combination of B, C, and E proteins, carpels (gynoecium) result from E and C protein functions, and finally ovules form due to C, D, and E proteins (Coen and Meyerowitz 1991; Moyroud and Glover 2017; Pelaz et al. 2000; Theißen 2001, Theißen 2016). Floral morphogenesis occurs after floral organ initiation, during which the size and shape of floral organs change. Finally, during maturation pigmentation, scent, and gustatory traits become fully developed (Shan et al. 2019). While the developmental trajectories of each type of floral organ differ, all involve proliferation, cell expansion and differentiation, and the establishment of lateral-medial, proximal-distal, and adaxial-abaxial polarities (Irish 2008; Moyroud et al. 2017; Sauret-Güeto et al. 2013; Walcher-Chevillet and Kramer 2016).

1.2.4 MYB transcription factors

Several developmental processes that occur during floral morphogenesis and maturation are regulated by MYB transcription factors. The MYB protein family is large and functionally diverse, occurring in all eukaryotes but selectively expanded in plants (Dubos et al. 2010). MYB proteins are characterised by a DNA-binding domain, called the MYB domain, that is generally comprised of up to four imperfect amino acid sequence repeats (R) that each form three alpha helices. A helix-turn-helix structure is formed from the second and third helices of each repeat. Different MYB protein classes vary in the number of adjacent repeats present (1R-MYB, 2R-MYB, 3R-MYB, 4R-MYB). The prototypic

MYB protein (c-MYB, animal cellular MYB) has three repeats named R1, R2, R3, and so repeats in other MYB proteins are named in accordance with their similarity to the c-MYB repeats. R2R3 MYB transcription factors, for example, are MYB proteins with two adjacent repeats within the MYB domain. The plant lineage has a very high diversity of MYB proteins, particularly R2R3-MYB transcription factors (Dubos et al. 2010; Katiyar et al. 2012; Martin and Paz-Ares 1997). MYB proteins containing two or more MYB repeats bind specific DNA sequence motifs cooperatively, thought to act like covalently linked dimers when interacting with DNA (Ogata et al. 1995). This dimerization may enable high affinity and specificity during protein-DNA interactions. MYB transcription factors seem to act both as immediate targets of other regulators and direct regulators of other genes, suggesting that they function at many levels in hierarchical regulatory networks (Dubos et al. 2010; Zheng et al. 2009). The bilateral symmetry of *A. majus* flowers develops due to ventralizing regulators (e.g. the MYB transcription factor DIVARICATA (DIV)), and dorsalizing regulators including CYCLOIDEA from the TCP family and the MYB RADIALIS (Almeida et al. 1997; Luo et al. 1995). AmMIXTA and the *Petunia* protein PhMYB1 are subgroup 9 R2R3 MYB transcription factors that induce epidermal cell outgrowths (Baumann et al. 2007; Glover et al. 1998; Noda et al. 1994). MYB transcription factors are also involved in the biosynthesis of flavones in all tissues and have a key role in pigmentation regulation (Stracke et al. 2007). MYB transcription factors have been associated with transcriptional regulation of betalains (Hatlestad et al. 2015), carotenoids (Sagawa et al. 2016), and are very well characterised in the regulation of anthocyanin pigmentation (Davies et al. 2012; Zhao and Tao 2015). The production of many phenylpropanoid volatiles, including benzaldehyde and eugenol, are also regulated by MYB proteins (Spitzer-Rimon et al. 2012). In *A. thaliana*, *Petunia hybrida*, and *Rosa* (variety Pariser Charme), PRODUCTION OF ANTHOCYANIN 1 (PAP1) functions in the regulation of both floral anthocyanin and volatile production (Zvi et al. 2012).

1.3 R2R3 MYBs in floral anthocyanin pigmentation

R2R3-MYB transcription factors regulate anthocyanin biosynthesis in many systems. They recognise *cis*-regulatory MYB-core and AC-rich elements (Franco-Zorrilla et al. 2014; Kelemen et al. 2015; Prouse and Campbell 2012), and different R2R3 MYB DNA-binding domains vary in DNA binding preferences, largely within these core motifs. R2R3-MYBs have a modular structure and most regions are highly variable; however, there is an activation or repression domain at the C-terminus and a highly conserved MYB domain at the N-terminus. Based on these conserved regions R2R3-MYBs have been categorised into subgroups, largely characterised from *Arabidopsis* (Stracke et al. 2001). Research in other angiosperm species has led to identification of additional subgroups and expansion of existing subgroups (Du et al. 2015; Jiang et al. 2004; Millard et al. 2019). R2R3-MYB proteins are involved in regulating many plant specific processes including cell identity and fate, stress responses, primary and secondary metabolism, and developmental processes (Dubos et al. 2010). Subgroups 5, 6, and 7 of the R2R3 MYB transcription factor family are associated with anthocyanin synthesis (Feller et al. 2011; Stracke et al. 2001).

1.3.1 The regulation of anthocyanin pigmentation

Flower colour is primarily formed through pigmentation with betalains, flavonoids, and carotenoids (Davies et al. 2012). Flavonoids produce the greatest variety of pigments, such as white or ivory flavones, flavonols, and flavanones; yellow chalcones and aurones; and anthocyanins. Anthocyanin is the floral pigment with the broadest distribution across flowering plants, responsible for red, pink, black, and blue colouration (Grotewold 2006). Anthocyanins accumulate in the vacuole and the petal hue they produce is, in part, dependent on vacuolar pH. The pH of the vacuole can affect the flavonoid

molecule redox state causing changes in the wavelengths of light that are absorbed (Zhao and Tao 2015). The amino acid phenylalanine is the precursor to anthocyanin synthesis, anthocyanins are produced from one branch of the flavonoid pathway. Phenylalanine is first converted to coumarate-CoA, which reacts with malonyl-CoA catalysed by chalcone synthase producing the naringenin chalcone. Using this product, downstream enzymatic reactions controlled by chalcone isomerase and,

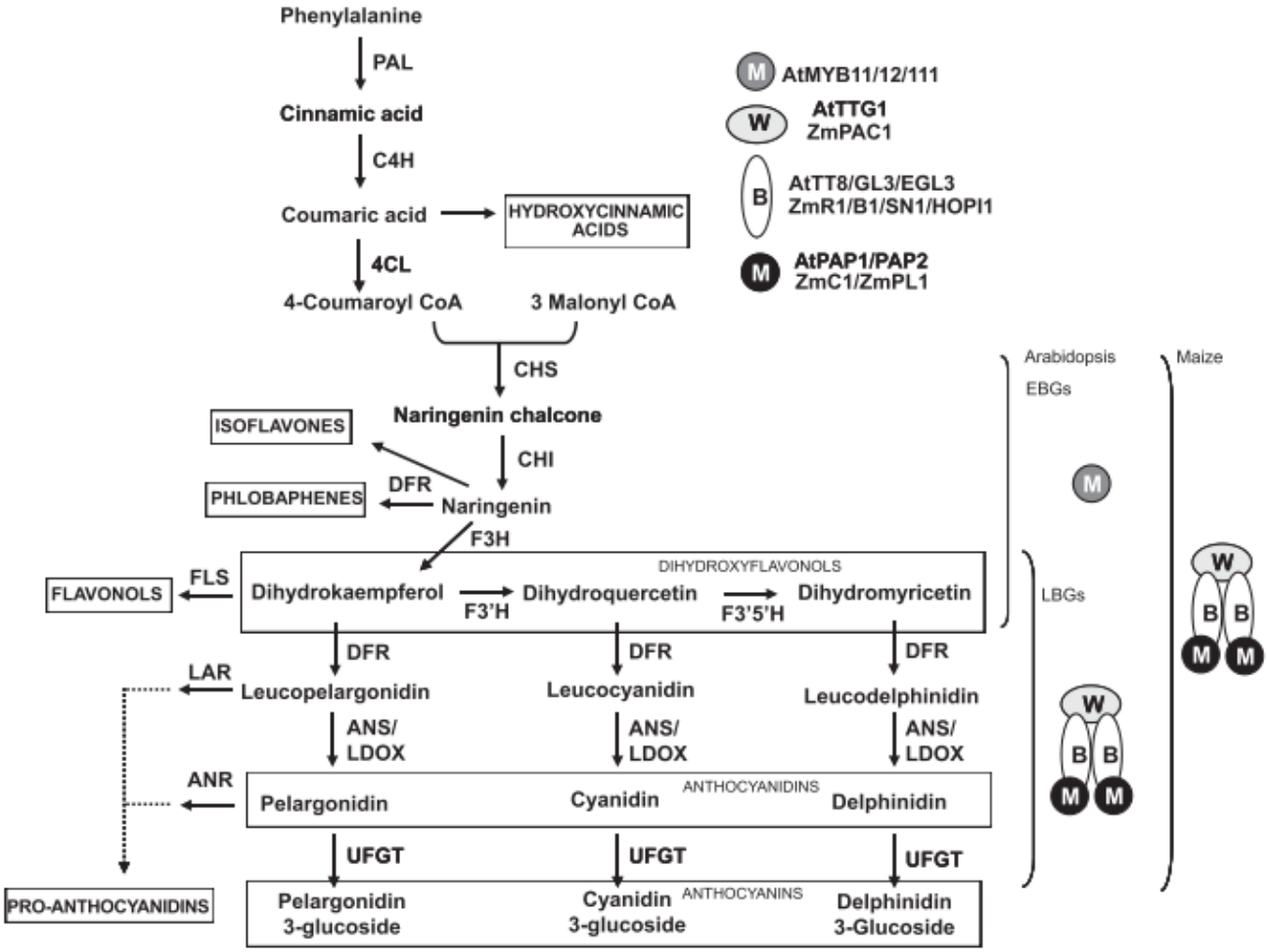


Figure 1.1. A simplified representation of the flavonoid pathway, denoting the phenylpropanoid pathway, providing precursors to the anthocyanin branch pathway (detailed) and other branches of the flavonoid pathway (end-products illustrated). Transcription factors controlling the *Arabidopsis* and maize anthocyanin pathway are illustrated: MYB (M), bHLH (B), and WD40 (W). PAL, phenylalanine ammonia lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4 coumarate CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3- hydroxylase; F3'H, flavanone 3'-hydroxylase; F3'5'H, flavanone 3'5'-hydroxylase; DFR, dihydroflavonol reductase; FLS, flavonol synthase; ANS/LDOX, anthocyanidin synthase/leucoanthocyanidin dioxygenase; UFGT, UDP-flavonoid glucosyl transferase; ANR, anthocyanidin reductase; LAR, anthocyanidin reductase. EBG, Early Biosynthesis Genes; LBG, Late Biosynthesis Genes. Reprinted from Petroni and Tonelli (2011) with permission from Elsevier.

subsequently, hydroxylases produce dihydroflavonols. The various dihydroflavonols formed are substrates for dihydroflavonol 4-reductase, which syntheses leucoanthocyanidins. Subsequently, anthocyanidin synthase converts these substrates into coloured anthocyanidins, for example cyanidin, pelargonidin and delphinidin (Fig 1.1). These anthocyanidins can be decorated by transferases, including acetylases and methyltransferases, and processed by 3-O-glycosyltransferases forming the chemically stable and water soluble anthocyanidin-3-O-glucosides (Chaves-Silva et al. 2018).

Anthocyanin compounds are broadly distributed across the plant lineage, and the anthocyanin pathway and its transcriptional regulation is well characterised in a diverse range of species (Grotewold, 2005; Hernández et al. 2009). Anthocyanin production is controlled by members of R2R3 MYB and basic helix-loop-helix (bHLH) families that form a complex (the MBW complex) with WR-repeat (WDR) proteins to activate anthocyanin biosynthesis pathway enzymes (Gonzalez et al. 2008; Ramsay and Glover, 2005). In many eudicot species different groups of transcriptional regulators control early and late biosynthesis genes, although partial overlap between regulators does occur. Comparing between species, transcriptional regulators can have differing affinities for various anthocyanin biosynthesis genes (Petroni and Tonelli 2011). Late biosynthesis genes are regulated by the MBW complex (Dubos et al. 2010; Ralf Stracke et al. 2007). MBW complex formation is also necessary in certain systems, including maize, to activate biosynthesis genes acting earlier in the anthocyanin pathway (Petroni and Tonelli 2011). Subgroup 6 R2R3-MYB transcription factors are strongly associated with regulation of late anthocyanin biosynthesis as part of MBW complexes in many systems and examples of anthocyanin phenotypes resulting from overexpression of these transcription factors are illustrated in Fig 1.2 (Albert et al. 2011; Lim et al. 2016). In petunia flowers anthocyanin synthesis is controlled by MBW complexes containing the WD40 AN11 and bHLH AN1 proteins that interact with two different subgroup 6 R2R3 MYB partners: AN2 in petals and AN4 in anthers (Petroni and Tonelli 2011; Quattrocchio et al. 2006; Schwinn et al. 2006; Spelt et al. 2000).

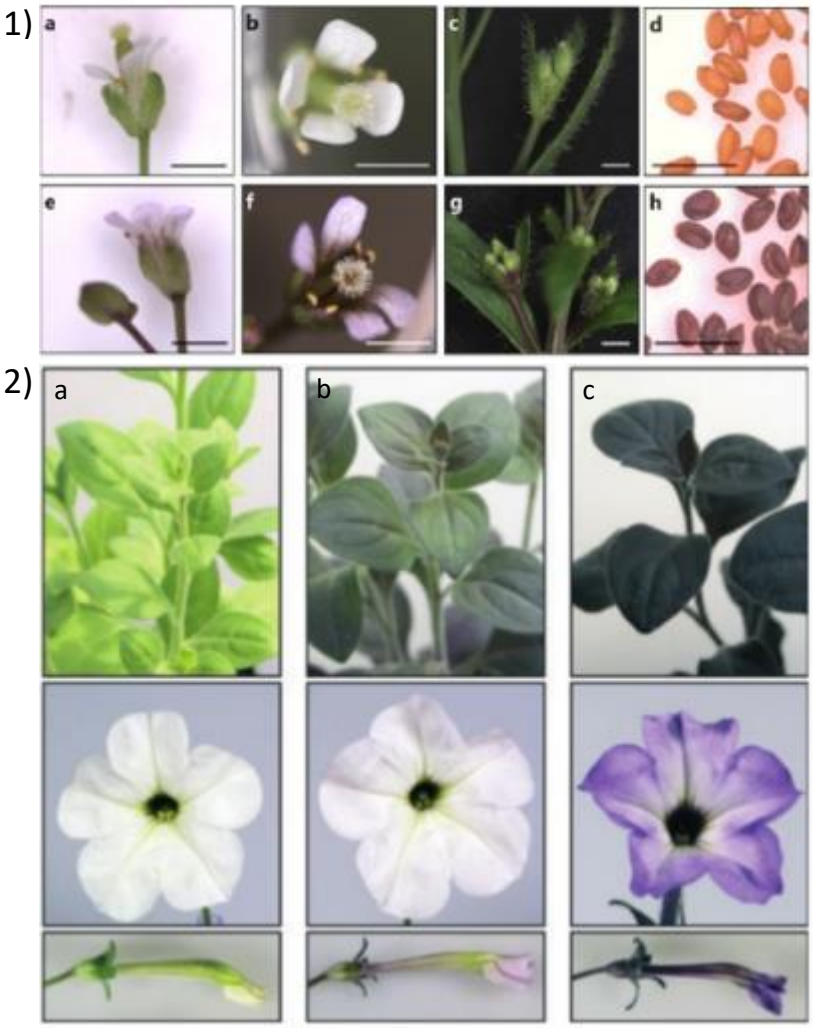


Figure 1.2. The induction of anthocyanin pigmentation from constitutive expression of subgroup 6 R2R3 MYB transcription factors 1) *Raphinus sativus* L. gene *RsMYB1* stably transformed into *Arabidopsis thaliana* (a-d) wildtype (e-h) transgenics. 2) *Petunia hybrida* genes *PURPLE HAZE (PHZ)* and *DEEP PURPLE (DPL)* stably transformed into *P. hybrida*. 2a) wild type 2b) *PHZ* transgenic line 2c) *DPL* transgenic line. Figures were reprinted by permission from 1) Springer Nature: Plant Cell Reports, Lim et al. (2016), 2) John Wiley and Sons: Plant Journal, Albert et al. (2011).

1.3.2 Floral patterning

Differences in pigmentation between cells within one floral organ or between floral organs can produce floral patterning including spots, stripes, and bicolour flowers (Davies et al. 2012; Gaskett, 2011; Leonard et al. 2011; Shang et al. 2011). Floral pattern formation requires genetic mechanisms that restrict the location of pigment. The flavonol and anthocyanin pathways have common precursors and in *M. lewisii* competition for these substrates leads to petal patterning (Fig 1.3). The R2R3 MYB transcription factor LAR1 represses anthocyanin production around the corolla throat, resulting in a white patch in an otherwise pink corolla – a trait potentially important for bumblebee pollination (Owen and Bradshaw 2011; Yuan et al. 2016). Within the white region of the petal, LAR1 activates flavonol synthase expression and this enzyme diverts dihydroflavonol substrates to flavonol (colourless or ivory) production and away from the anthocyanin pathway (Yuan et al. 2016). In *Clarkia*

gracilis spatial restriction of dihydroflavonol-4-reductase (DFR) (an anthocyanin synthesis enzyme) is a result of DFR activation by a spatially restricted R2R3 MYB transcription factor (CgMyb1). *Cis*-regulatory differences between the promoter sequences of each *CgMYB1* allele causes each allele to be expressed in different petal regions resulting in variation in corolla spot position (Martins et al. 2013; Martins et al. 2017). The production of petal patterns requires adjacent cells to adopt distinct fates during development and an anthocyanin regulatory network model has been developed using *Petunia* data to explore this. Incorporated into the model was a putative R2R3 MYB and information from previous research on the role of *Petunia* MBW complexes and R3 MYB repressors in floral pigmentation patterning. Cell-specific anthocyanin production was presented as a result of coordination between activator and repressor proteins that interact through a series of feedback loops (Albert et al. 2014). Similarly, an R2R3 MYB activator and an R3-MYB repressor identified in *Mimulus* regulate petal anthocyanin spots, with a mode of action compatible with (but not confirmed as) the reaction-diffusion model or classic Turing instability (Fig 1.3) (Ding et al. 2018; Turing 1953). The reaction-diffusion model demonstrates how spatial patterns in tissues can develop through a self-activating activator and a repressor protein interacting, with the latter able to inhibit the activator along a diffusion gradient (Meinhardt and Gierer 2000). Some floral phenotypes are highly complex requiring coordinated regulation of multiple pigment pathways and different cell types, such as those involved in the formation of sexually deceptive petal spots in *Gorteria diffusa* (Thomas et al. 2009). Characterising the mechanisms that control complex patterning will prove challenging but will be aided by further establishment of model systems to address these questions.

1.4 Floral pigmentation and species diversification

The discovery of heritable phenotypes and use of novel traits for developmental innovations are key components of adaptation (Specht and Howarth 2015). To understand the way in which these traits evolve requires not only a developmental perspective but also consideration of the population genetics and specific environmental context within which the traits emerge because selection acts upon the phenotype (Fernández-Mazuecos and Glover 2017). Coupling insight into genetic architecture of specific traits with understanding of how causal mutations originate and spread in populations enables more thorough understanding of natural variation and how it can lead to speciation (Nunes et al. 2013). Studies of colouration may provide a particularly good model for investigating these different elements of microevolution. Adaptive colouration is an important contributor toward plant and animal fitness, and a lot is known about the genetic control of colour traits, with mechanisms beginning to be elucidated. Insights into the evolution of pigmentation in systems such as the monkeyflower *Mimulus* and *Heliconius* butterflies have provided understanding of the interplay between ecology, evolution and development (Orteu and Jiggins 2020; Wu et al. 2008; Yuan 2019). Therefore, developing additional flower model systems could provide both information relating to floral trait genetics and a predictive framework that contributes toward broader developmental evolutionary understanding (Stern 2011 in Nunes et al. 2013).

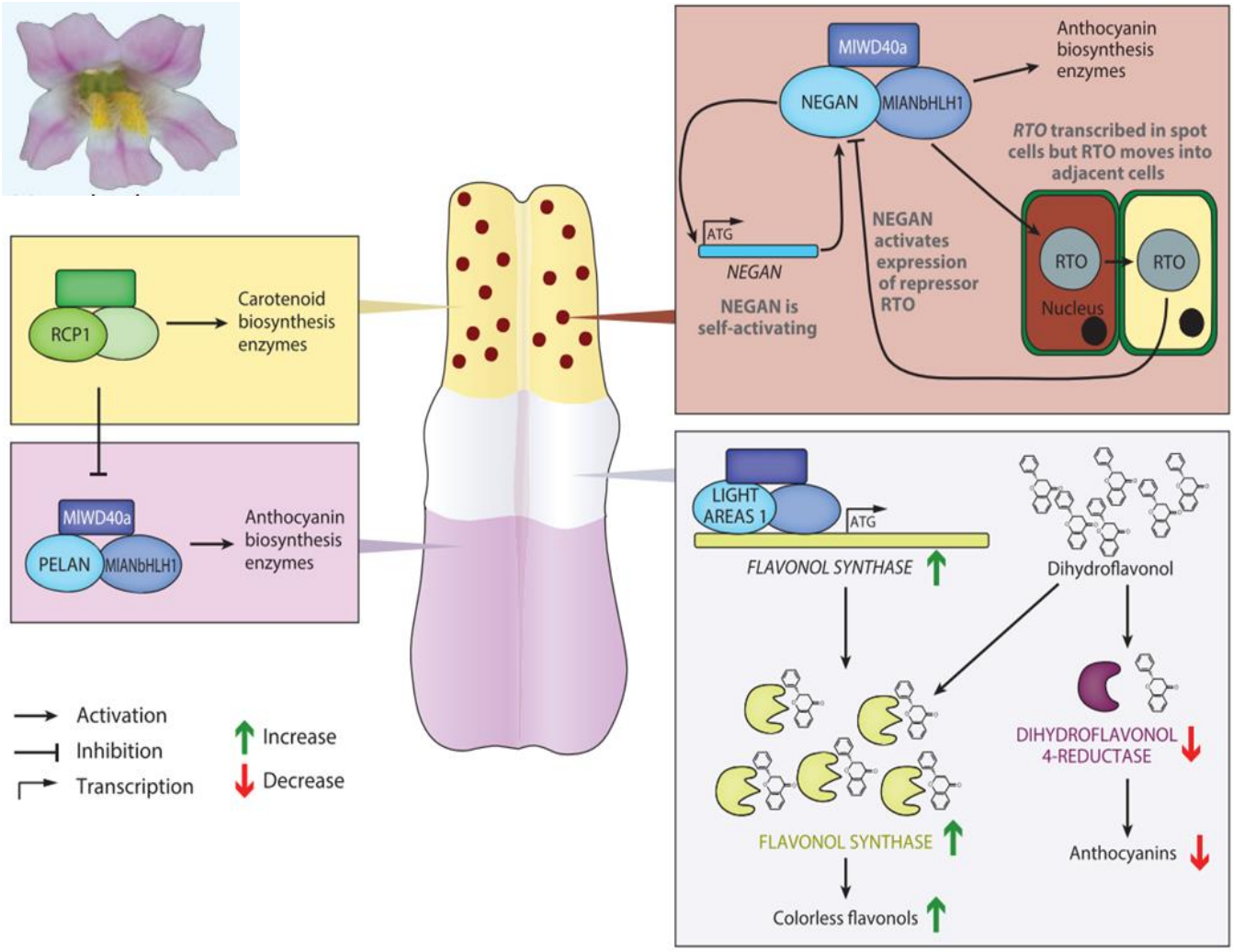


Figure 1.3. Our current understanding of the genetic processes underlying pigmentation traits in *Mimulus lewisii* (depicted in the top left) petals are illustrated for (clockwise from the right) anthocyanin spots (Ding et al. 2018), reduction in anthocyanin in the white region of the corolla throat (Yuan et al. 2016), petal lobe anthocyanin pigmentation (Yuan et al. 2014), and nectar guide carotenoid pigmentation (Sagawa et al. 2016). Black arrows indicate regulation of a gene/protein or synthesis of a product, green and red arrows indicate a relative increase or decrease in a particular substrate/enzyme/expression of a gene/product. The regulatory genes listed are R2R3 MYBs. Reproduced with permission by Annual Reviews, Fattorini and Glover (2020).

1.5 The study system *Gorteria diffusa*

1.5.1 Growth habit and phylogenetic placement

Gorteria diffusa is an herbaceous and self-incompatible daisy (Asteraceae), described by Thunberg in the 18th century. The name *G. diffusa* refers to the plant's diffuse growth habit; multiple branches emerge radially from the rootstock and grow prostrate along the ground (Duncan and Ellis 2011; Ellis and Johnson 2010). Plants take on average three to four months to flower and have approximately 20-60 open inflorescences (Duncan and Ellis 2011; Stångberg et al. 2013). Predominantly an annual species, there are perennial coastal populations (Ellis and Johnson 2009). Germination within *G.*

diffusa is dependent on rain and suitable weather conditions, only a few achenes may germinate while the others are dormant for one to several seasons. *G. diffusa* belongs to the subtribe Gorteriinae in the largely southern African tribe Arctotoideae (Funk and Chan 2008; Karis 2006). Roessler (1959, 1973) split *Gorteria* into three species: *G. corymbosa*, *G. diffusa*, and *G. personata*. Recent phylogenetic studies on Arctotideae - Gorterrinae demonstrated that Roessler's taxonomic concepts are not compatible with the evolutionary history of the group. While *Gazania* constitutes one clade, *Gorteria* and *Hirpicium* species are intermingled in two additional clades (Stångberg et al. 2013). The suggested revised taxonomy *sensu* Stångberg et al. (2013) is illustrated in Fig 1.4.

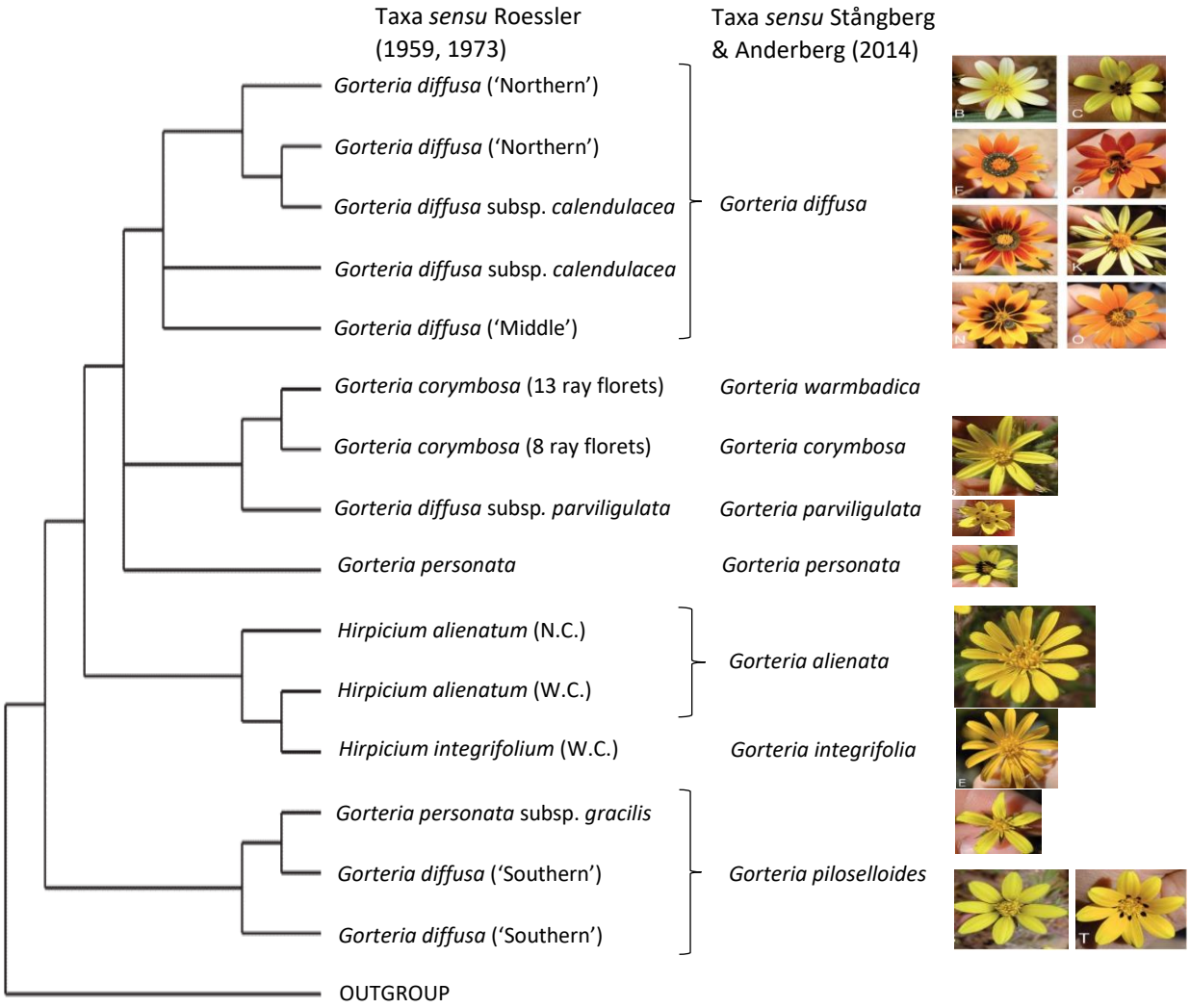


Figure 1.4. Simplified phylogeny of *Gorteria*. Taxa sensu Roessler (1959, 1973) are denoted in the left-hand column and taxa *sensu* Stångberg & Anderberg (2014) are the right-hand column. 'Northern', 'Middle' and 'Southern' describe the part of the *G. diffusa* range that samples were collected from. N.C. is 'Northern Cape' and W.C. is 'Western Cape'. Photographs depict the species they are adjacent to, and the relative size of each capitulum is approximately to scale. Modified from Stångberg and Anderberg (2014) and photographs taken with permission by the author and publisher (International Association for Plant Taxonomy) from Stångberg et al. (2013).

1.5.2 Habitat and geographical distribution

G. diffusa inhabits the winter rainfall zone of southern Africa, with a distribution spanning southern Namibia to the Richtersveld and Namaqualand to the Western Cape of South Africa (Fig 1.5) (Duncan and Ellis 2011; Roessler 1959). This semi-arid Succulent Karoo biome is considered relatively young, as most endemic lineages originated less than 10mya. The Succulent Karoo is a biodiversity hotspot (de Jager and Ellis 2017; Myers et al. 2000) and the Namaqualand subregion contains approximately 38000 angiosperm species and >400 species of Asteraceae in 55000km^2 (Snijman 2013). Aridification that started in the Miocene is often cited as a potential pressure that increased diversification in this biome (Diester-Haass et al. 2002; Linder 2003; Tyson and Partridge 2000; Verboom et al. 2014). Pollinators are another potential driver of angiosperm speciation in the Succulent Karoo, as plant species tend to have strong pollinator specialisation that may contribute toward reproductive isolation (Ellis et al. 2014; Johnson 2010; Linder 2003). Pollinator interactions may be particularly important because Namaqualand has a narrow temporal flowering window and high occurrence of self-incompatible annual species (Kemp et al. 2019; de Waal et al. 2014). Each spring, *G. diffusa* contributes to mass flowering displays, dominated by daisies, that characterise this region (Kemp et al. 2019). *G. diffusa* grows in large colonies, from late July to early October, in sand or clay soil, on flats and rocky hill slopes (Duncan and Ellis 2011). Abundant floral visitors to Namaqualand plant communities include flies (Bombyliidae, Tabanidae), Hopliini beetles (Scarabacidae), and bees (Apoidea) (Ellis and Johnson 2009; Struck 1994). The most prevalent daisy flower colour present varies geographically, for example the majority of daisies in upland areas are orange and white colouration is dominant in coastal plains. Changes in pollinator community composition occur over small distances and this variability in pollinator assemblage is consistent with a role in promoting these clustered assemblies of flower colour patterns. The dominant pollinators in each community investigated (flies: *Megapalpus capensis* and *Rhigioglossa* sp. respectively) interacted most strongly with the floral colour patterns that were overrepresented. This is consistent with species coexistence being driven by facilitation, or evolutionary convergence, of shared colour patterns preferred by locally dominant pollinators (Kemp et al. 2019). The repeated evolution of different colour patterns in multiple distantly related daisy lineages suggest that flower colour in Namaqualand daisies may be evolutionarily labile. This is also evident in intraspecific differences in floral colour patterns that occur allopatrically within some species, including *G. diffusa* (Kemp et al. 2019).

1.5.3 Asteraceae floral architecture

The Asteraceae capitulum is a compressed inflorescence consisting of many flowers on a receptacle surrounded by bracts (Bello et al. 2013; Weberling 1992). It contains two types of flowers: actinomorphic disc florets are centrally positioned, and at the periphery of this central zone are zygomorphic ray florets. Ray florets have long ventral petals that are fused together and reduced dorsal petals (Bello et al. 2013; Garcês et al. 2016). The capitulum functions as an effective reproductive unit, with disc florets specialised for reproductive processes and ray florets functioning in pollinator attraction. This morphology provides 'a more flexible basis for breeding system evolution than does a single flower' (Lane, 1996 and Jeffrey, 2009 in Bello et al. 2013). In *G. diffusa* actinomorphic disc florets are male in the centre and male or hermaphroditic towards the periphery of the capitulum. The disc florets mature acropetally and so before the central florets open the outer disc florets are often pollinated, and the capitulum wilts (Johnson and Midgley 1997; Thomas et al. 2009). *G. diffusa* ray florets are sterile and attract pollinators using a ligule of four fused petals that is

brightly coloured. Dark petal spots develop across this ray floret corolla (Karis 2007; Thomas et al. 2009).

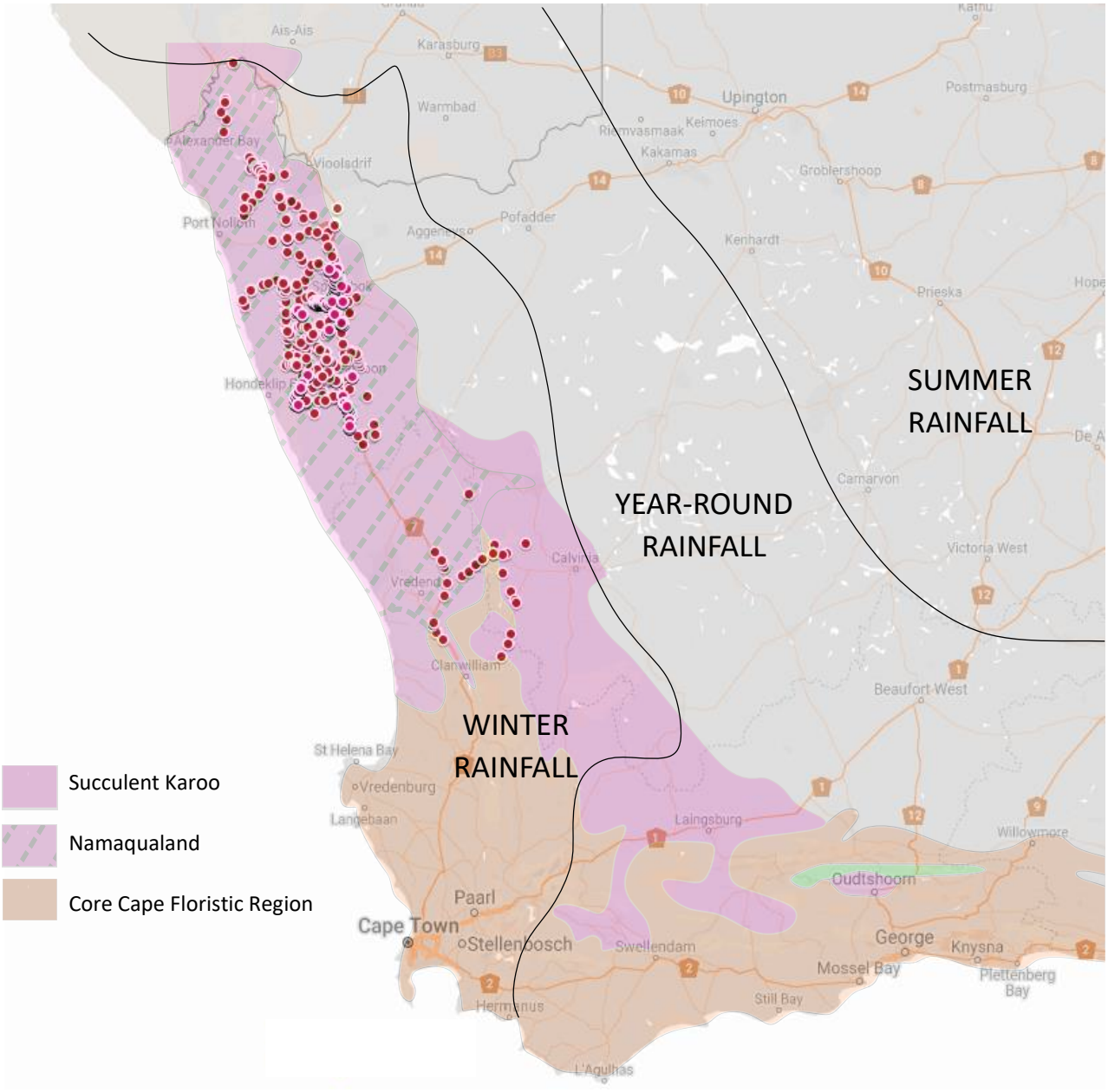


Figure 1.5. Map of the West coast of South Africa. Black lines indicate approximate boundaries between different rainfall regions indicated on the map as winter, year-round, and summer rainfall based on Chase and Meadows (2007). Coloured overlay depicting different regions are based on those outlined in (Cowling 2015). Points indicate locations where *G. diffusa* has been sampled – this is not an exhaustive representation. Map was taken from google maps.

1.5.4 Intraspecific variation

G. diffusa floral phenotype is highly variable between populations across its distribution (Thomas et al. 2009). Groups of geographically proximal populations were observed to have similar phenotypes, with sharp boundaries where floral traits varied dramatically in adjacent populations. Based on this geographical variation in *G. diffusa* inflorescence phenotype, populations were assigned to different

floral morphotypes (by Allan Ellis and Steve Johnson) named after the areas in which they were found (Fig 1.6). A quantitative investigation across the species range was then conducted, focussing particularly on whether clusters of discrete geographical floral forms occur or whether floral variation is continuous. 240 populations were located across succulent Karoo vegetation, by driving along public roads and stopping every 20-30 km – covering an area of 35000 km^2. Capitulum trait measurements, including spot structural traits, were taken from 5 - 7 plants in each of 53 populations and used in cluster analyses. If floral forms were discrete, multiple populations of a morphotype should have combined trait variation levels equivalent to that within-populations. Clusters did comprise several populations of the same floral form, demonstrating that capitulum traits differ between floral morphotypes – with a few exceptions including Okiep and Kleinzee, and the spot traits of Rich and Kleinzee. A caveat of cluster analysis is that it always finds discrete groups, but population trait scores also revealed discrete clusters in multidimensional scaling space; for inflorescence and spot traits, five and eight clusters were found, respectively. Single clusters always contained all individuals from a population and several geographically adjacent populations, with the exception of the Okiep - Kz complex (Ellis and Johnson 2009). There is variability in floral phenotype within morphotypes, for example, variation is relatively high in Okiep populations (Ellis and Johnson 2009) and Stein contains both spotted and non-spotted individuals (personal observation, Allan Ellis observation). Within other floral forms, such as Cal, floral phenotype is highly consistent. It is evident that the *G. diffusa* species complex contains several geographically discrete floral forms identifiable by differences in capitulum phenotype.

Phenotypic comparisons between greenhouse grown individuals and wild plants found no significant differences in floral trait measurements for all morphotypes tested, except an increase in flower size in greenhouse populations (Ellis and Johnson, 2009; Thomas et al. 2009; personal observation). This demonstrates that the specific phenotypes of each floral form are heritable. Contact regions occur between many morphotypes and usually there is a steep clinal transition between floral forms. Individuals with intermediate phenotypes are found in some of these contact zones (Ellis and Johnson 2012). DNA samples from transects spanning two adjacent floral morphotypes, and the corresponding contact zone, are being used to determine how *G. diffusa* genetic structure is organised (by Boris Delahaie). This should provide insight into reproductive isolation and patterns of introgression in the system. The species has undergone recent taxonomic revision, as *G. diffusa* was found to be polyphyletic and the most Southerly morphotype (Worcester) is now considered a separate species. The presence of three well supported and phenotypically distinct groups prompted the suggested taxonomic classifications illustrated in Fig 1.4. While seven morphotypes from the 'Northern' distribution area formed an unresolved polytomy, the most Northerly morphotypes Khubus and Rich formed a sister group to this clade (bootstrap=89%, PP=1). Nieuw, from the 'Middle' distribution area, was resolved as a sister group to both of these clades but with low support (bootstrap=74%, PP=0.87). Ultimately, to determine the appropriate taxonomic status of each *G. diffusa* floral morphotype will require a well-resolved molecular phylogeny of the species (currently in progress), crossing experiments, and rigorous examination of floral and vegetative trait variation (Ellis and Johnson 2009). The genetic similarities between morphotypes, despite stark floral phenotypic differences, imply potential recent and rapid floral evolution (Stångberg et al. 2013). As such, once morphotype taxonomic status is resolved, *G. diffusa* will provide an excellent system for investigating the evolution of ecologically relevant floral traits in a species potentially undergoing incipient speciation.

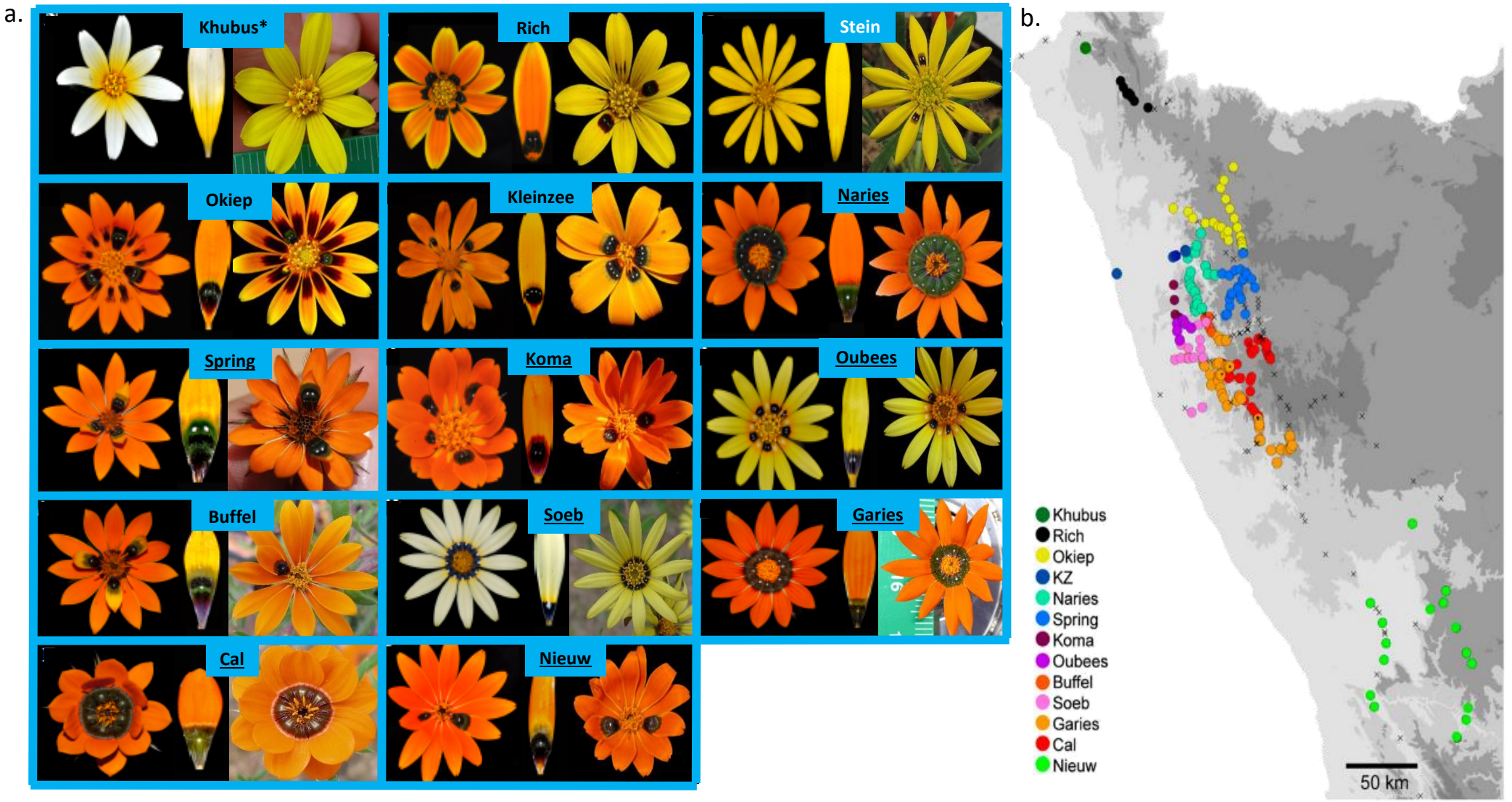


Figure 1.6. The morphotypes of *G. diffusa*. a) Two photographs of the capitula of each morphotype, with the morphotype name given in the blue box. * indicates some individuals produce simple black spots at the base of the ray florets, but this is not depicted in the images provided. Where text is black these morphotypes are included in the phenotypic characterisation of Ellis and Johnson (2009) white text indicates more recently identified morphotypes not included in this analysis. Underlined morphotypes were included in the quantitative characterisation of spot phenotypes in Ellis et al. (2014). b) Map of Namaqualand in South Africa indicating the distribution of 13 of the *G. diffusa* floral morphotypes. Points represent the occurrence of *G. diffusa* colour-coded by morphotype and crosses are sites where *Gorteria* did not occur. Photographs in a) were taken by Allan Ellis, Róisín Fattorini, and Matthew Dorling. Figure b) was reproduced with permission by John Wiley and Sons from Ellis and Johnson (2009).

1.5.5 Petal spots

Ellis and Johnson (2009) found that the inflorescence traits contributing most to differentiation between floral forms included spotted ray floret number, presence of raised spot florets, and presence of a black capitulum ring. These criteria complement the loose categorisation of *G. diffusa* spot types by Stångberg et al. (2013): (1) Simple dark basal spots on all ray florets; (2) Raised spots typically present on 1-5 ray florets within a capitulum – these tend to be more structurally complex, comprised of multiple epidermal cell types (Thomas et al. 2009); (3) Curvature of the ray floret lamina resulting in spots with a raised appearance, which are present on all ray florets in a capitulum. These (3) spots are generally structurally complex, although without the epidermal protrusions (papillae) sometimes present in (2). Spot types are not all mutually exclusive, with some capitula having both simple dark basal spots and raised spots on 1-5 ray florets. The number of raised spots per capitulum can vary within a plant and, in some morphotypes, spots can be completely absent from some individuals within a population. Spot types and variation are illustrated in Fig 1.7. Preliminary evidence suggests that spot presence can also vary within plants of the morphotype Stein, but this needs to be verified with a larger sample size using individuals from every Stein population.

The morphotype Nieuw was used in a detailed morphological analysis of complex raised petal spots. Nieuw petal spots were described as 'deeply textured and richly coloured, such that the overall appearance is a shimmery, almost iridescent elaboration of the petal' (Fig 1.8) (Thomas et al. 2009). This intricate phenotype is due to three specialised cell types that have distinct size, shape, and pigmentation - compared to unspecialised pavement epidermal cells of the ray floret. Central highlight cells have a smooth cuticle and outer cell wall, they are relatively small, have little pigmentation (appearing white), and reflect UV. Interior cells are relatively short and round, with pigmentation that is variable and appears to be cell autonomous. Petal tissue composed of interior cells appears green and textured, undulating around smooth highlight cell tissue. Multicellular papillae are iridescent groups of swollen epidermal cells containing high concentrations of anthocyanin; these structures form an arc around the periphery of the spot. Papillae are absent in several morphotypes and papillae shape and elaboration also vary between *G. diffusa* morphotypes. Other differences in complex spots between morphotypes include the number of highlights present, and the arrangement of spot cell types relative to one another (Thomas et al. 2009). As the spot develops at the base of four congenitally fused petal lobes, specialised cell types span all or several lobes. Spot patterning develops within a zygomorphic flower, across a subsection of the corolla, but is only present on 1-4 ray florets. As such, spot development in Nieuw is complex over several morphological scales: at the cellular level (spot), considering the whole floral organ (ray floret), and in overall inflorescence development (capitulum) (Thomas et al. 2009). While simple basal spots are found in other *Gorteria* species and South African Asteraceae genera, the morphology of Nieuw spots is unusual. *Gazania* and *Arctotis* also contain species developing intricate petal spots, but the complex patterning is mostly caused by pigment accumulation. This also applies to non-daisy taxa such as *Pelargonium* (Geraniaceae) and *Rhododendron* (Ericaceae). It is the substantial elaborations and modification of the epidermis in *G. diffusa* petal spots that makes them highly unusual morphologically. The presence of spots on only a subset of ray florets is also unique amongst the aforementioned spotted species (Thomas et al. 2009).

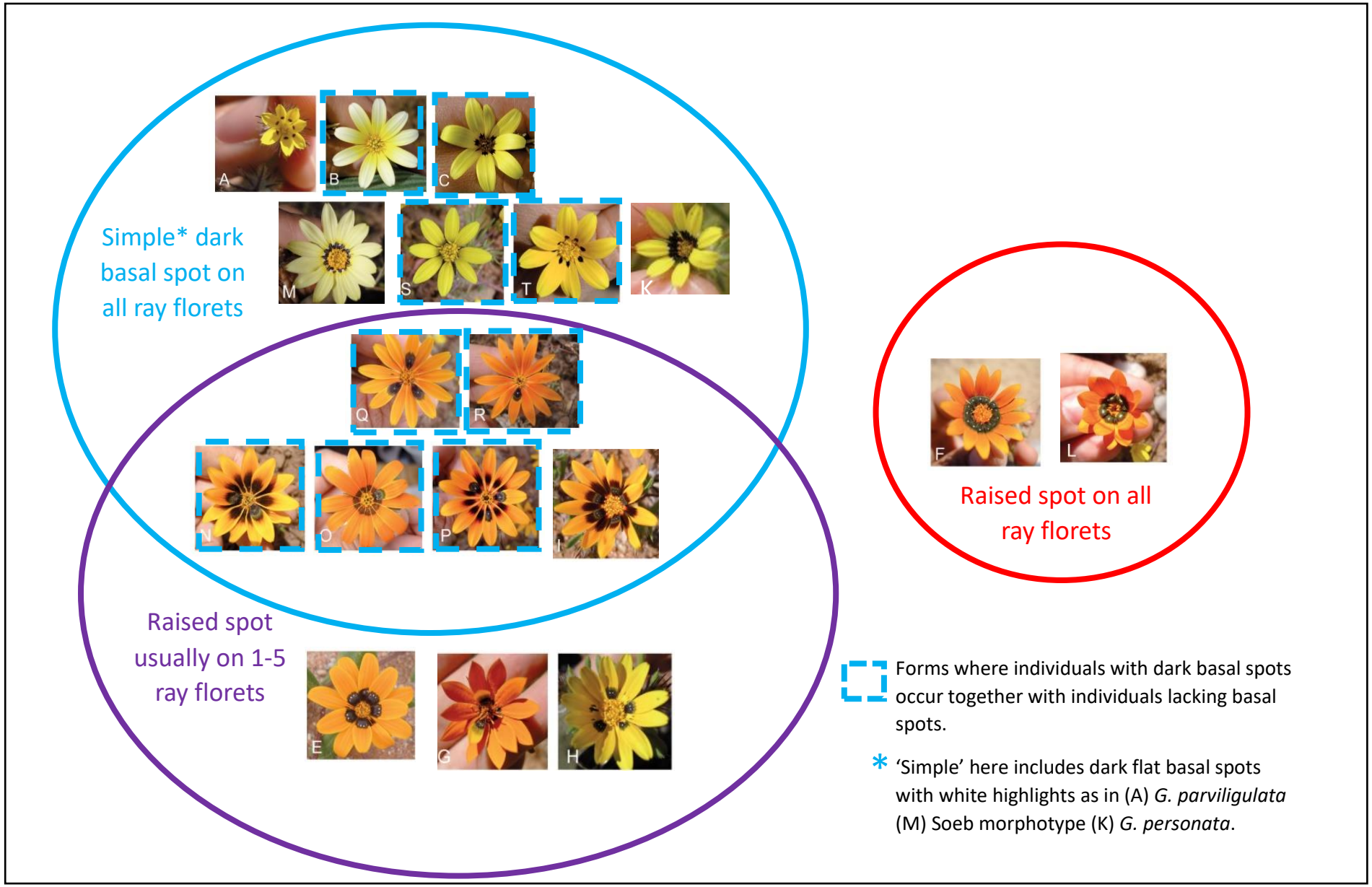


Figure 1.7. Graphical representation of the petal spot types in *Gorteria* (as described by Stångberg and Anderberg, 2014) and the dynamic complexity within the system. This is not an exhaustive description of the variation within *Gorteria*, for example, there are also some non-raised spots that are not present on all ray florets. All images are *G. diffusa* as defined in Stångberg et al. 2013 except for (A) *G. parviligulata* and (K) *G. personata.* Photographs were taken with permission by the author and publisher (International Association for Plant Taxonomy) from Stångberg et al. (2013).

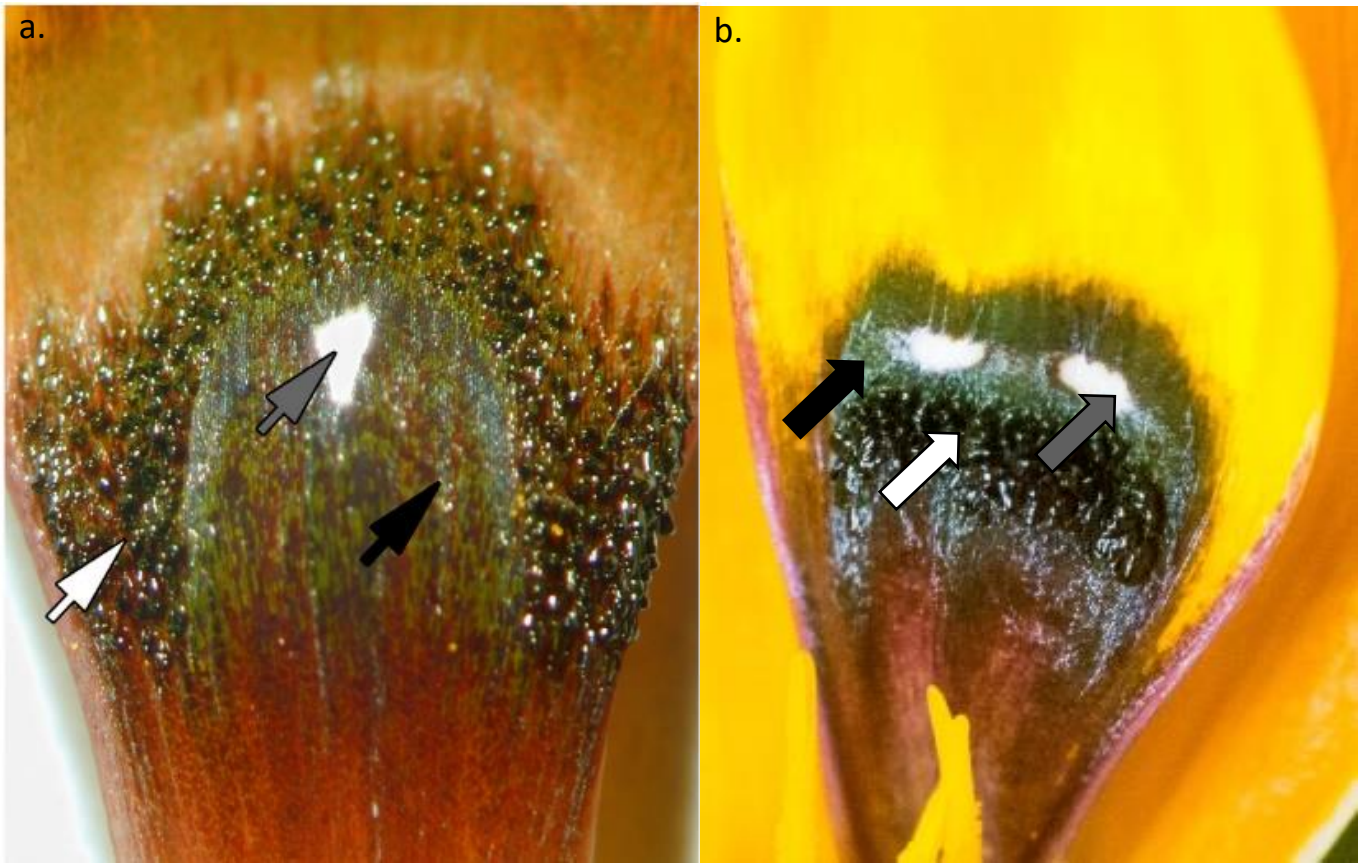


Figure 1.8. G. diffusa complex petal spots in two morphotypes. The petal spot epidermis is rich in anthocyanin and composed of three specialised cell types: green interior cells (black arrow), multicellular papillae (white arrow), and white highlight cells (grey arrow). a) A light micrograph of the Nieuw morphotype, reproduced with permission by John Wiley and Sons from Thomas et al. (2009). b) A photograph of the Spring morphotype (taken by Boris Delahaie).

1.5.6 Petal spot evolution and pollinators

G. diffusa petal spots attract the Bombyliidae bee fly *Megapalpus capensis* (Wiedermann). This is a widespread and abundant pollinator in the Succulent Karoo, it feeds on the nectar and pollen of many angiosperm species, including *G. diffusa*, and is considered a dominant pollinator of multiple daisy genera (Ellis and Johnson 2009; de Jager and Ellis 2013; Johnson and Midgley 1997). *M. capensis* pollinates taxa with convergent dark spots in the capitulum, or dark reproductive organs; implying this bee-fly may have contributed toward the evolution of the 'dark spotted' pollination guild in Namaqualand (de Jager and Ellis 2012; Johnson 2010). To compare pollinator species across floral forms, *G. diffusa* capitula were inspected in 62 populations. Floral visitors were identified and checked for the presence of *G. diffusa* pollen. *M. capensis* visited all 10 floral morphotypes surveyed, bees were found collecting pollen on northern floral forms, and horse flies (Tabanidae) frequented three morphotypes and were occasionally found on three additional floral forms. The visitation rates of *M. capensis* were mostly higher than those of other pollinators, but Soeb had more tabanid *Rhigiolossa* spp visitors. *M. capensis* flies always made contact with stigmas during floral interactions and all inspected flies were carrying *Gorteria* pollen, suggesting *M. capensis* is an effective *G. diffusa* pollinator (Ellis and Johnson 2009).

M. capensis exhibits three different behaviours on *G. diffusa* capitula: pollen and nectar feeding, brief inspection visits, and active attempts to copulate with the spot (Fig 1.9) (Ellis and Johnson 2010;

Johnson and Midgley 1997). This is one of only two cases of pollination by sexual deception reported in a family other than Orchidaceae, with the other being *Iris paradoxa* (Vereecken et al. 2012). The petal spots are thought to mimic females, as only male flies attempt copulation (Ellis and Johnson 2010; Johnson and Midgley 1997). This pseudocopulatory response is only observed on specific morphotypes with raised complex spots on a subset of ray florets (Spring, Buffels, and Nieuw) (Ellis and Johnson 2010; Thomas et al. 2009). Capitula with simple spots on each ray floret elicited feeding behaviours only, while male inspection behaviours occurred on more elaborate spots. Spot models investigating how visual, tactile, and olfactory components contribute to *M. capensis* attraction found contrasting preferences between male and female flies, for example, only males had a preference for spots with papillae (de Jager and Ellis 2012). As floral odour extract did not trigger male copulation behaviour, it is unlikely to contain gender-specific pheromonal signals contributing to the sexual deception. However, it is possible that not all relevant compounds were extracted, and males did show preference for odour extracts of spotted ray florets in the model combining all sensory components. Therefore, *G. diffusa* may be using multimodal signals to deceive males, activating different fly sensory systems to improve signal detection (Candolin 2003; de Jager and Ellis 2012).

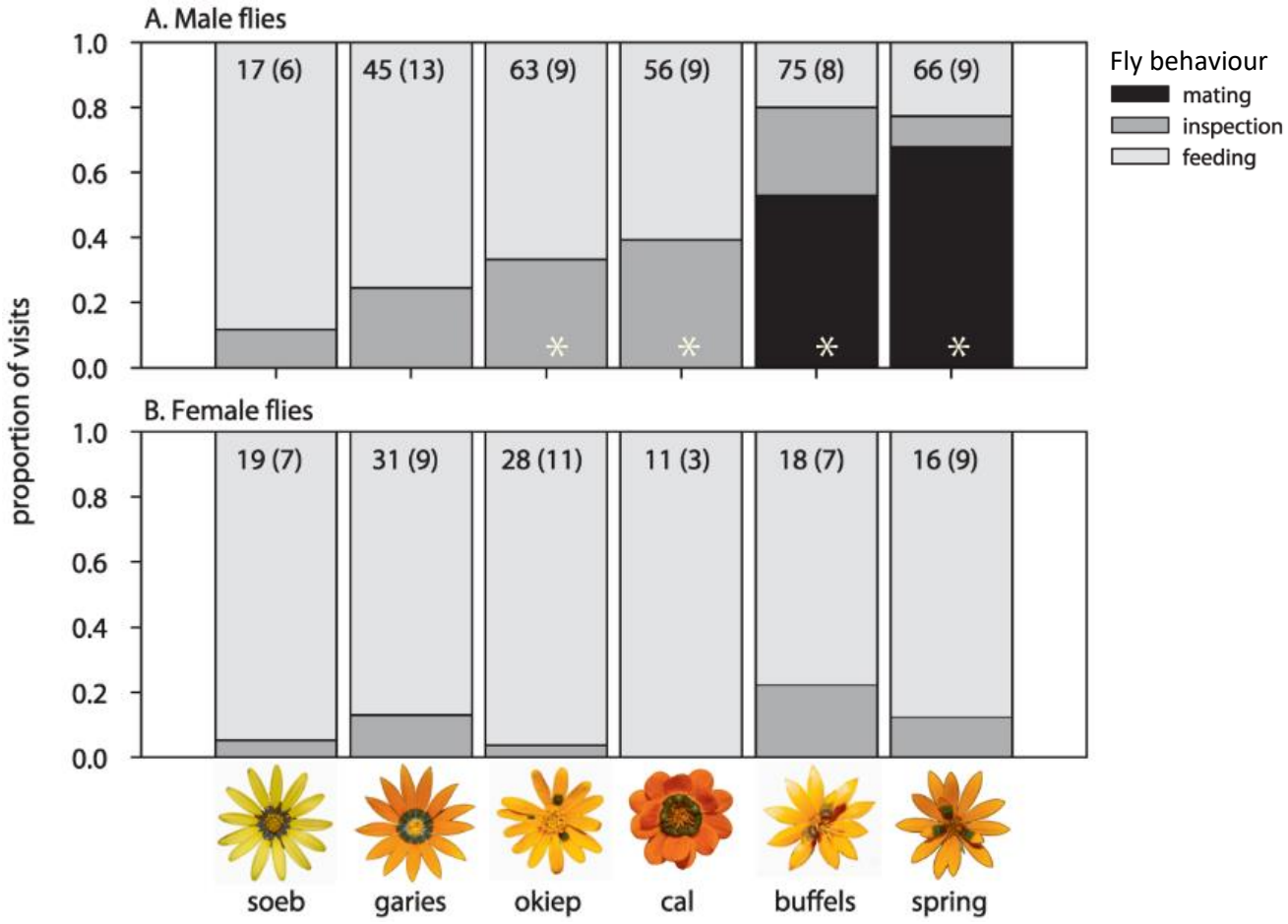


Figure 1.9. The behavioural responses that *Megapalpus capensis* flies exhibit upon visitation to *G. diffusa* floral morphotypes (A) male flies (B) female flies. Behaviour is indicated by bar colour: light grey (feeding), dark grey (inspection), black (mating). Data were collected during controlled cage experiments. The number of observed visits were listed and the number of flies used in the experiment are given in parentheses. Asterisks indicate floral morphotypes where male and female flies exhibited combinations of behavioural responses that were significantly different. This figure was reproduced with permission from Ellis and Johnson (2010).

As mate-seeking flies are more active than feeding flies, this behaviour may be advantageous for the plant by enabling more effective pollen transfer between individuals and so increasing outcrossing rates (Ellis and Johnson 2010). Fly mimicry may have evolved in *G. diffusa* through antagonistic coevolution, as deceived males learn to avoid female-mimicking spots over time (de Jager and Ellis 2014). In this scenario, sexual deception would impose a reproductive cost on male *M. capensis* and this would be selected against through learned discrimination between real females and petal spots. In turn, the plant is under selection to retain the pollinator to maximise reproductive success, so the spot phenotype becomes more deceptive – reducing the ability of the fly to detect the deception. It is evident that the high intraspecific variation within *G. diffusa* evolved without pollinator shifts. However, contrasting fly behaviours and pollinator environments between floral forms could still, theoretically, have contributed to divergence within this species. Contrasting preferences for spot phenotype between male and female *M. capensis* could impose divergent selection on *G. diffusa*. If the fly sex ratios varied between populations, but remained stable over generations, selection exerted by the most abundant gender could determine floral phenotype (de Jager and Ellis 2012). Geographically variable, but stable, variation in the effectiveness of each gender as pollinators could also promote floral divergence. Flies do not have preferences for *G. diffusa* floral forms based on the plant community context or innate preferences, so assortative visitation based on floral phenotype is unlikely to be a factor in maintaining floral morphotypes at contact regions (Ellis and Johnson 2012).

So far investigations into how *M. capensis* may have contributed to *G. diffusa* floral divergence have yielded no clear explanations. Floral variation may also have evolved due to selection from other types of floral visitors such as other pollinators, herbivores, and seed predators (Galen 1999, Gómez 2003, Rey et al. 2006). Additional potential barriers to gene flow between morphotypes include ecological isolation, postzygotic isolation mechanisms, and physical geographical barriers. The latter is possible for certain floral forms, as boundaries between them and adjacent morphotypes occur at landscape features, such as rivers, that may reduce gene flow. The isolated *G. diffusa* populations could then diverge through the neutral process of genetic drift (Ellis and Johnson 2009). Potential prezygotic and postzygotic barriers to gene flow are currently being investigated.

1.5.7 Current research into *G. diffusa*

A phylogeny resolving the evolutionary relationships between *G. diffusa* morphotypes is currently underway (Mellers 2016; Delahaie unpublished). The genetic processes underlying phenotypic differentiation between morphotypes within *G. diffusa* are being investigated through analyses of genetic structure across morphotype hybrid zones, coupled with ecological research to establish potential intraspecific barriers to gene flow. Candidate genes that may contribute toward key aspects of petal spot development have been identified from a recent transcriptome, these include genes that may regulate papillae development (Kellenberger unpublished). The developmental aspect of this project follows on from previous work by Mellers (2016) and Walker (2012). The previous work established that *G. diffusa* petal ray florets are pigmented by the anthocyanin cyanidin in the subset of morphotypes investigated (Thomas et al. 2009, Walker 2012). A *G. diffusa* MYB gene (*GdMYB8*) was isolated from a Spring transcriptome analysis as a potential petal spot anthocyanin regulator (Walker 2012). This gene clustered within subgroup 6 of the R2R3 MYB transcription factors and could induce anthocyanin production in a heterologous host (Mellers 2016).

1.5.8 Project objectives

The overall project objectives were two-fold: to further our understanding of the genes regulating *G. diffusa* petal spot development and to provide a first insight into *G. diffusa* population genetic structure. Investigations were primarily conducted within a single morphotype (Spring) to provide a comprehensive knowledge base which can be used in future as a key resource for between-morphotype comparative research. Spring was chosen as the focal morphotype because it has been used for previous investigations conducted within the laboratory and Spring petal spots are particularly interesting regarding functionality, as they are sexually deceptive to male bee fly pollinators.

Investigating the spatial scale of gene flow across the Spring morphotype range

During a field season in the Northern Cape of South Africa, several plants were sampled within multiple populations across the Spring morphotype range. Floral phenotypic measurements were recorded, and leaf samples taken. DNA was extracted from these samples and used in a genotyping by sequencing analysis to investigate population genetic structure. The data were analysed using a number of complementary approaches and the relationship between genetic differentiation and geographical distance between sites was investigated. We hypothesised that the spatial scale of gene flow may be relatively small, due to limited seed dispersal. Phenotypic data was collected to enable a comparison of genetic structure with any phenotypic differentiation present. This is the first population genetics study conducted within *G. diffusa*. The results will contribute toward understanding of the genetic structure of species within the Succulent Karoo biodiversity hotspot.

Characterising anthocyanin petal spot regulators

Several candidate R2R3 MYB transcription factors were identified as potential petal spot anthocyanin regulators, homologous to the *GdMYB8* gene previously identified. The expression patterns of these genes were determined across three *G. diffusa* morphotypes (Spring, Cal, and Stein) using qRT-PCR. Genes that were upregulated in spotted petal tissue were stably transformed into *N. tabacum* to assess whether these genes were sufficient to induce ectopic anthocyanin production in a heterologous system. The expression patterns of the transgenes and *N. tabacum* anthocyanin synthesis enzymes were determined through qRT-PCR to provide insight into the functioning of the transgenes within *N. tabacum*. Potential downstream targets in the anthocyanin synthesis pathway were characterised and expression patterns investigated to determine the contribution of these genes to pigmenting petal spots.

Investigating the function of regulatory proteins within G. diffusa

Attempts were made to further develop a *G. diffusa* stable transformation protocol to enable overexpression of *GdMYB8* genes within *G. diffusa*. Biochemical assays were used to determine whether GdMYB8 proteins could bind to promoter motifs of candidate anthocyanin synthesis enzymes. The assays conducted were yeast one-hybrid experiments, enabling assessment of whether GdMYB8 proteins could bind to promoter fragments and activate gene transcription within yeast, and gel shift assays which are currently ongoing. Promoter regions of *GdMYB8* genes were isolated through genome walking as a resource for identifying candidate *MYB8* regulators once an efficient *G. diffusa* stable transformation has been developed.

2. General methods

Methods common to multiple data chapters are listed here, while more specific procedures are given within the relevant data chapter.

2.1 Plant growth conditions

(Relevant for all data chapters)

Seeds of *Gorteria diffusa* were collected from wild populations in Namaqualand, South Africa between June - September 2013 (morphotype Spring collected by Greg Mellers), 2016 (multiple morphotypes collected by Alan Ellis), and 2018 (multiple morphotypes collected by Boris Delahaie, Róisín Fattorini, Jurene Kemp). Whole fruits were soaked in water for approximately 16 hrs before being sown in a mixture of 20% sand and 80% Levington's M2 potting compost. The plants collected were grown in either (1) Plant Sciences Department greenhouse kept at 21°C, with a photoperiod of 16 hrs and approximately 60% humidity, (2) Cambridge University Botanic Garden greenhouse at 16 - 33°C, a photoperiod of 16 hrs, and ambient humidity or (3) Cambridge University Plant Growth Facility (16 hrs light/8 hrs dark, 20°C, 60% humidity). The following morphotypes were grown: Spring, Cal, Stein.

2.2 RNA and DNA extractions

(Relevant for all data chapters)

2.2.1 Tissue preparation for DNA and RNA extraction

Tissue were harvested for DNA/ RNA extraction and immediately flash frozen in liquid nitrogen. The plants used for tissue harvesting were > 4 weeks old and 100mg of tissue was taken, either young leaves (approx. 0.5 – 4cm in length), whole flower buds, or developing ray florets were used in the extractions (Sections 4.2, 5.2, 6.2). While frozen, the tissue was pulverised either using a sterile pestle and mortar or using a tissue lyser (Qiagen Tissuelyser II). For the latter, a glass bead (5mm diameter) was added to the microcentrifuge tube containing the sample and the tubes were shaken at 30Hz for 20 - 30 secs.

2.2.2 Extraction of genomic DNA

DNA was extracted from powdered leaves using cetylmethylammonium bromide (CTAB). 500µl of CTAB buffer (Appendix 1) were added to approximately 100mg of leaf powder, vortexed, and incubated at 55°C for 1 hr. 5µl of RNase A were added and samples were incubated for 30 mins at 37°C. Each sample was mixed with 500µl chloroform: iso-amyl alcohol (24:1), vortexed, and centrifuged (10 mins at 12000xg). The aqueous phase was retained, the volume was estimated, and it was mixed with an equal volume of chloroform: iso-amyl alcohol (24:1). Again, the solution was vortexed, centrifuged, and the aqueous phase was retained. Based on the estimated volume of the aqueous phase, 0.08 volumes of ammonium acetate and 0.58 volumes of isopropanol were added. Samples were kept at -20°C for 1 hr – 16 hrs. Following centrifugation for 15 mins at 12000xg, the resulting pellet was washed with 70% (v/v) ethanol and 100% (v/v) ethanol before resuspension in 20µl of sterile water. DNA extracts were stored at -20°C. Whether or not the extraction was successful was determined by nanodrop readings and visualisation of the DNA on an agarose gel (Sections 2.3.6 and 2.3.8).

Extraction of gDNA for genotyping was done using a quicker method. 250µl of extraction buffer (Appendix 1) were added to the approximately 15mg of leaf material, which was ground with a pestle

and then vortexed and centrifuged (1 min at 12000xg). The supernatant was extracted and mixed with 200µl of isopropanol, the solution was incubated at room temperature for 2 mins before centrifugation (5 mins at 12000xg). The resulting pellet was washed with 70% (v/v) ethanol before resuspension in 30-50µl of TE buffer (Appendix 1).

2.2.3 CTAB RNA extraction

20 - 100mg of frozen ground tissue were added to 700µl of CTAB buffer (Appendix 1) and vortexed. This solution was mixed, through vortexing, with 700µl of chloroform: isoamyl alcohol (24:1) and incubated at 55°C for 15 mins. Following centrifugation for 15 mins at 10,000rpm (at 4°C) the upper phase was transferred to a clean tube. Once again, chloroform: isoamyl alcohol was added – an equal volume to the upper phase removed in the last step; the mixture was vortexed, centrifuged (10,000rpm, 15 mins, 4°C) and the upper phase was removed to a clean tube. 0.33 of the volume of the upper phase was calculated and this amount of chilled 8M lithium chloride was added to precipitate the DNA, mixed by pipetting, and incubated at 4°C overnight. This solution was then centrifuged for 20 mins at 10000rpm (at 4°C) to pellet the RNA. The supernatant was removed and kept for later DNA precipitation. 150µl of 3M sodium acetate and 500µl of 100% (v/v) ethanol were added to the RNA pellet before centrifugation for 20 mins at 13000rpm (at 4°C). The supernatant was removed, 700µl of 70% (v/v) ethanol were added, and the pellet was centrifuged for 20 mins at 13000rpm (at 4°C). The supernatant was discarded, and the RNA pellet was left to air-dry before resuspension in 25µl of sterile deionised water.

2.2.4 Removing DNA contamination from RNA

Turbo DNA-free kit

Due to the time taken to collect plant material for qPCR, small quantities were obtained for RNA extraction. As such, TURBO DNA-*free* kit (Ambion) was used to maximise the final concentration of RNA obtained. The final reaction volume was 25µl - 28µl depending on the starting concentration of RNA and was made up to this volume using sterile deionised water. TURBO DNase buffer (0.1 volumes of 10x buffer) and 1µl of TURBO DNase were added to the RNA and mixed by pipetting before incubation at 37°C for 30 mins. DNase inactivation reagent (0.1 volumes) were added and mixed by pipetting. The solution was incubated at room temperature for 5 mins and then centrifuged at 10000rpm for 1.5 mins. 22µl of the supernatant, containing the RNA, were transferred to a clean tube and stored at -80°C.

Incubation with DNase I

Deoxyribonuclease I (thermofisher) was added to RNA samples to remove any contamination by genomic DNA. 10µg of RNA were treated in a 100µl reaction containing 1µl (1U) of DNase I and 1x concentration of buffer (Appendix 1) and incubated at 37°C for 1 hr. The DNase was then removed using the phenol-chloroform clean-up detailed below.

Phenol-chloroform clean-up of RNA

A phenol-chloroform clean-up of the RNA was used to remove the DNase enzyme from the previous step and purify the RNA. The reaction volume was made up to 500µl using sterile deionised water and 500µl of cold 28:24:1 phenol:chloroform:isoamyl alcohol were added, and the solution was vortexed. Following centrifugation at 14000rpm for 5 mins (at 4°C), the upper phase was transferred to a clean

tube. To precipitate the RNA 750µL of 95% (v/v) ethanol 5% (v/v) sodium acetate (pH 5.5) were added, mixed by pipetting, and incubated at -20°C for an hour. The solution was centrifuged for 5 mins at 14000rpm (at 4°C) to pellet the RNA. The supernatant was discarded and 500µl of 70% (v/v) ethanol were added, before centrifugation for 5 mins at 13000rpm. Following removal of the supernatant the pellet was left to air-dry and then resuspended in 20µl of sterile deionised water.

Checking for gDNA contamination

Following procedures to remove DNA contamination and purify RNA samples, PCR reactions were conducted to ensure DNA contamination was not present at detectable levels. RNA samples were diluted 3 - 5 fold in ddH_2O and used as templates in PCRs with GdActin primers (Appendix 2), along with positive gDNA controls and a negative control (no template). 15µl - 20µl of the PCR product was run on an agarose gel. As RNA is not amplified by PCR, DNase treatment was considered successful if no DNA band was present on the gel in RNA lanes. If a band was produced through PCR amplification of an RNA sample, DNase treatment was repeated, and the RNA sample checked for DNA contamination through PCR and visualisation on a gel again.

First-strand cDNA synthesis

To synthesise cDNA for PCR, reverse transcription of RNA was performed using BioScript reverse transcriptase (Bioline). 1µl of Oligo dT primer (Invitrogen) was mixed with 1-4µg of DNase treated RNA and heated for 5 mins at 70°C. The Oligo dT primer complements and binds to the poly-adenine tail of the mRNA molecules present. 1µl 40mM dNTP, 4µl 5X reaction buffer, 1µl BioScript (50U), and 1µl RNase inhibitor were added, mixed, and the solution was incubated for 40 mins at 42˚C and then 85˚C for 5 mins to denature the reverse transcriptase. The cDNA was stored at -20°C. In cases where the 3' end of a particular transcript was unknown and needed to be amplified, GeneRacer Oligo dT primer (Invitrogen) was used in the reaction rather than Oligo dT primer. The GeneRacer Oligo dT primer also complements and binds to the poly-adenine RNA tail but adds an additional known sequence (sequence 1) to the 3' end of the gene. In subsequent PCR reactions, a reverse primer specific to sequence 1 was coupled with a forward primer in the known gene sequence, enabling amplification of the end of the gene and the 3'UTR. To synthesise cDNA for qRT-PCR SuperScript III reverse transcriptase was used. 400-1000µg RNA were mixed with 1µl of 40mM dNTP and 1µl of 50µM Oligo dT primer. The reaction volume was made up to 13µl with sterile deionised water and incubated for 5 mins at 65°C. The samples were left on ice for 5 mins before being mixed with 4µl of 5X first strand buffer, 1µl of 0.1M DTT, RNase OUT (40U) (ThermoScientific) and SuperScript III reverse transcriptase (200U). The reaction mix was incubated for 1 hr at 50°C and then heated for 15 mins at 70°C to denature the enzymes. The cDNA was stored at -20°C. To assess whether cDNA synthesis was successful, the cDNA was used as a template in PCRs conducted using primers that amplify the *GdActin* (*G. diffusa*) or *NtActin* (*N. tabacum*) – depending on the cDNA species of origin.

2.3 Isolating DNA sequences

(*Relevant for Chapters 4, 5, and 6)*

2.3.1 Primer design

Previous and current work on *G. diffusa* generated comparative transcriptomes from spotted and non-spotted ray florets (Walker 2012, Kellenberger unpublished), and gene hunting through PCR was previously undertaken by Greg Mellers (Mellers 2016). These sequences were used as a basis for

primer design to isolate specific genes. Unless otherwise stated, primers were designed to be 18 - 30bp long, with GC content 40 - 60%, and primer pairs had melting temperatures within 5°C of each other. All primers were ordered via Integrated DNA Technologies (www.eu.idtdna.com).

2.3.2 Polymerase chain reaction (PCR)

A generalised PCR programme and master mix solutions are outlined in Appendix 2, but this protocol was modified when applying PCR to specific experiments. PCR was carried out by the following PCR machines: TECHNE TC-4112, Techne 3Prime, and Labnet Multigene OptiMax Gradient Cycler. Agarose gel visualisation was used to determine the success of PCR reactions (Section 2.3.6).

2.3.3 Genome walking (inverse PCR)

Modified from the Ren et al. 2005 method

Following CTAB gDNA extraction from the morphotype Spring, pairs of restriction enzymes that produced compatible sticky ends were used to digest the *G. diffusa* genome. The compatible sticky ends of the gDNA fragments annealed forming circular DNA – some of which contained the target gene sequence and adjacent genomic regions. The following pairs of restriction enzymes were used in gDNA digestion to create genomic libraries: Bgl II/ Bam HI, Pst I/ Nsi I, Pst I/ Shf I, Sal I/ Xho I, and Eco RI/ Mfe I. 50U of each enzyme was used to digest 10µg of gDNA in a 500µl reaction with the corresponding buffer, found using NEB double-digest-finder (https://nebcloner.neb.com/#!/redigest). This solution was incubated for 16 hrs in a 37°C oven. Following digestion, the DNA solution was purified using a phenol-chloroform method and the resultant DNA pellet was washed in 70% (v/v) ethanol, left to air-dry, and resuspended in 11µl of sterile H_2O. 500ng of this purified DNA were ligated at 16°C for 16 hrs in a 50µl reaction mix containing 2µl T4 ligase (NEB), 5µl T4 ligase buffer, and sterile H_2O. Following circularisation of the digested gDNA, this DNA was used as a template in PCR reactions. Primers and nested primers (detailed in relevant Chapter) were designed from the known gene sequence to amplify outwards into regions flanking the gene. Two consecutive PCRs were conducted using PCR Bio (PCRBIOSYSTEMS). The PCR product from the first reaction was diluted 1:50 and used as a template in the second PCR reaction using nested primers.

Using Universal GenomeWalker 2.0

Following CTAB gDNA extraction, 4 gDNA libraries were created using Dra I, Eco RV, Pvu II, and Stu I restriction enzymes as detailed in Universal GenoeWalker 2.0 User Manual (https://www.takarabio.com/assets/documents/User%20Manual/Universal%20GenomeWalker%202.0%20User%20Manual_040314.pdf). Briefly, genomic DNA was used to produce adaptor-ligated genomic DNA fragments, by digesting the gDNA with restriction enzymes and then ligating this to the GenomeWalker Adaptor provided. These genomic DNA libraries were then used as templates for PCR and nested PCR – using forward primers that annealed to the adaptor and reverse primers designed from the gene of interest. The PCR enzyme PHUSION (NEB) was used (as per the manufacturer's instructions) and PCR product from the first reaction was diluted 1:50 and used as a template in the subsequent nested PCR. The nested PCR product was run on a gel and bands were gel extracted (outlined in Section 2.3.7) and sequenced. Alternatively, the PCR product was first cloned and then sent for sequencing.

2.3.4 Degenerate primers

Degenerate primers were designed to amplify the full length of genes, when either the beginning or end was unknown, and to try to characterise 5' regions. The *G. diffusa* portion of the gene available from either a transcriptome (Walker 2012), or previous gene hunting (Mellers 2016) was BLASTed using nucleotide BLAST on NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against Asteraceae sequences (Asteraceae (taxid:4210)). These sequences were downloaded and aligned in Geneious (https://www.geneious.com). Using these alignments, sections of sequences that were relatively conserved were selected to design degenerate primers from. Within these selected regions SNPs were categorised into those that vary between 2 nucleotides (e.g. in all sequences that base pair position is either an A or a C) and those that vary by 4 (i.e. all bases are present at the position across sequences). Primers were designed so that multiplication of the variable sites present did not surpass 64, for example, if 2 sites varied by 2 nucleotides and 2 sites varied by 4 nucleotides 2*2*4*4 = 64. Once the sequence sections were determined for each primer the correct degenerate base symbols were found using IUPAC code.

2.3.5 Rapid amplification of cDNA ends (3'RACE)

Amplification and sequencing of 3' ends of cDNA fragments was undertaken using a 3'RACE protocol. As described in Section 2.2.4, cDNA was synthesised with a known adaptor sequence of DNA added to the 3' end of every strand. Primers specific to the adaptor sequence were used in a nested pair of PCR reactions with gene specific primers – enabling amplification of unknown 3' mRNA regions. The initial PCR product was diluted 1:50 with ddH_2O and this was used as the template in the nested PCR reaction.

2.3.6 Agarose gel electrophoresis

The success of nucleic acid extraction and amplification was determined using agarose gel visualisation. Agarose was dissolved in 0.5x TBE buffer (Appendix 1) by heating to achieve a final concentration of 1-1.5% (w/v) and ethidium bromide (EtBr: 0.1µg/ml) was added. The gel was set in a gel cassette and nucleic acid samples were mixed with gel loading buffer (Appendix 1) and loaded onto the gel. A current of 110V was passed through the gel, separating DNA fragments according to size. DNA bands were visualised using UV light (Syngene G:Box), as EtBr fluoresces upon binding with nucleic acid. The bands were compared to a DNA ladder (Hyperladder 1Kb Bioline) containing bands of known size, which was run on the gel simultaneously with the samples.

2.3.7 Agarose gel extraction and purification

To isolate DNA fragments of the correct length, relevant bands were excised from the gel under a UV light using a razor blade. DNA was purified using either a Monarch DNA Extraction Kit (NEB #T1020), a Clontech NucleoSpin PCR Clean-Up and gel extraction kit, or an Invitrogen Purelink quick gel extraction kit according to the manufacturer's guidelines. Extraction and purification involved the insertion of solubilised gel into a column and cleaning with an ethanol solution. Subsequently, the purified DNA was eluted from the column membrane in 20µl of sterile water.

2.3.8 Quantifying DNA and RNA concentration

The concentration of DNA and RNA was quantified using a NanoDrop 2000 UV-Vis spectrophotometer. Measurements were made relative to a blank sample composed of the solute in which the RNA/ DNA

was resuspended and stored. The absorbance for DNA was quantified at 260nm and for RNA at 280nm. Purity of the samples was assessed by checking the absorbance graph for a smooth curve and the absorbance ratios at 260/280nm and 260/230nm. Low values indicate the possibility of protein contamination and salt/ solvent contamination, respectively. In combination with this, agarose gel visualisation was used to provide an approximation of DNA/ RNA quantity by comparing the brightness of the DNA band to that of a ladder band of known quantity.

2.4 Cloning and sequencing

(Relevant for Chapters 4, 5, and 6)

2.4.1 Producing chemically competent *E. coli* DH5α

Chemically competent *E. coli* DH5α cells were produced in the laboratory. DH5α are used for cloning because they recombine at a low rate, which is advantageous for plasmid replication. DH5α cells from a glycerol stock were grown for approximately 16 hrs in a 37°C incubator. 4ml of this culture were inoculated in 120ml of LB (Appendix 1) and incubated at 37°C for 3 hrs. The cells were pelleted by centrifugation for 5 mins at 1844xg, resuspended in 10ml of magnesium chloride (100mM concentration), and kept on ice for 5 mins. The cells were pelleted through centrifugation (as described above) and resuspended in 2ml of *E. coli* freezing solution (Appendix 1). Cells were aliquoted, frozen in liquid nitrogen, and stored at -80°C.

2.4.2 PCR product ligation into a plasmid

For cloning purposes, DNA fragments amplified through PCR were ligated into plasmids. Two vector systems were employed for cloning: pGEM – T Easy (Promega Corp.) or pBluescript SK(-) vector cut in house with EcoRV. pBluescript SK(-) vectors were prepared by digesting the plasmid with EcoRV, cleaning the digested plasmid using phenol-chloroform (detailed in Section 2.2.4), and resuspending the resulting pellet in 50µl of TE. Ligation into the pGEM vector occurs through annealing of 'sticky ends' – extended regions of single stranded deoxyadenosine nucleotides produced from polymerases that terminate with adenine (e.g. PCR biotaq), these regions overlap with the pGEM vector complementary single stranded deoxythymidine. In PCR reactions using polymerases that produce blunt ends (e.g. Phusion DNA polymerase), PCR products were ligated into pBluescript SK(-) vector. 5µl ligation reactions included T4 DNA ligase, 0.5µl T4 DNA ligase buffer, 0.5µl 10mM dNTPS, and 0.5-3.5µl DNA template (dependent on DNA concentration). Ligation reactions were left at 16°C for 16 hrs.

2.4.3 Transformation of plasmids into *E. coli* and plasmid purification

25µl of chemically competent DH5α cells (Section 2.6.1) were added to 5µl of DNA ligated into vectors (Section 2.6.2), mixed, and left on ice for 10 mins. The mixture was heated at 42°C for 1 min 10 secs, promoting the uptake of the vector, containing the DNA fragment of interest, by competent cells. Subsequently, cells were left on ice for 5 mins. 700µl of LB liquid media (Appendix 1) was added, solutions were mixed, and incubated at 37°C for 30 mins – 1 hr. Prior to this, LB agar plates were prepared using LB agar medium (Appendix 1) containing 100mg/l ampicillin to select for antibiotic resistance. 100mM IPTG and 20µl X-Gal (50mg/ml) were added to the surface of the set agar plates to enable recognition of colonies containing the transgene through blue-white selection (explained below). Following incubation of the competent cells in liquid LB, 100µl of cell suspension were plated onto LB agar plates. The remaining cell suspension was centrifuged at 5000xg for 10 mins, 500µl of

supernatant were removed, and the pellet was resuspended in the remaining liquid and plated. Once dry, plates were incubated at 37°C for 16 hrs to allow colony growth.

Colonies that were potentially successfully transformed with the desired DNA fragment were white as opposed to blue. An antibiotic resistance gene within the vector ensured that only *E. coli* colonies transformed with the vector successfully grew on the media. Colony colouration indicated whether the vector contained a foreign DNA insert, as vector cloning sites are within the *lacZ* operon - which functions in lactose metabolism (encoding *β*-galactosidase). IPTG present in the medium induces the *lac* operon, while X-gal is a *β*-galactosidase enzyme substrate. If the operon is interrupted by a foreign gene insert, the X-gal will not be digested by *β*-galactosidase and so colonies appear white as opposed to blue resulting from X-gal digestion. Individual colonies were sampled using a sterile cocktail stick, which was then shaken into 25µl of sterile water. Subsequent incubation for 5 mins at 95°C lysed bacterial cells and 5µl of this solution were used as a template in a PCR reaction to determine vector insert size. Primers were used in colony PCR reactions (Appendix 2, unless otherwise stated in methods), the enzyme eco-Taq DNA polymerase (homemade) was used in these reactions and master mix components are listed in Appendix 1. Successfully transformed colonies were identifiable through gel electrophoresis, as having a product length of the gene insertion plus an additional 230-260bp of vector. Once identified, cells from these colonies were cultured in 3ml LB and 100mg/l ampicillin for 16 hrs at 37°C.

Desired plasmids were extracted and purified using an alkaline lysis method. Following overnight growth, cultures were pelleted by centrifugation at 1600xg for 1 min and all LB was removed. 300µl of cold Solution 1 (Appendix 1) and 5µl RNase A were added and the solution was vortexed. 300µl of Solution 2 (Appendix 1) and 300µl of Solution 3 (Appendix 1) were added and mixed by inversion. Solutions were left on ice for 5 mins and then centrifuged for 10 mins at 16000xg. The supernatant was removed, retained, and mixed with 640µl of isopropanol. Following centrifugation, the DNA pellet was washed with 70% (v/v) ethanol before being left to air dry and resuspended in 20µl of sterile water. DNA concentration of purified plasmid was determined using a nanodrop 2000 (ThermoFisher) spectrophotometer.

2.4.4 Sanger sequencing

Sanger sequencing was completed by the Department of Biochemistry, University of Cambridge.

2.5 Vector creation

(Relevant for Chapters 4, 5, and 6)

2.5.1 Gibson assembly

The Gibson assembly method was used to produce DNA fragments encoding individual MYB8 proteins with attached histidine tags and to ligate this into an existing vector (pETM11 Appendix 3). PCRs were conducted to create 2 bricks of DNA for use in the Gibson assembly reaction. Each brick was created by primers designed to amplify the desired DNA fragment, while adding additional base pairs identical to a section of the second target DNA fragment at either end. As extended regions of each DNA brick overlapped with one another, bricks could subsequently be annealed and ligated into a single vector during the Gibson assembly reaction. Following PCR, gel extraction and gel purification (outlined in Section 2.3.7) were conducted, and the DNA concentration of each brick was determined using the nanodrop. The volume of each solution required to give a concentration of 50ng of DNA brick was

calculated and used as a template in the Gibson assembly reaction – a 20µl reaction (4µl 5x isothermal buffer, 0.08µl T5 exonuclease (1U), 0.25µl Phusion DNA polymerase (2U), 2µl Taq DNA ligase (40U), ddH2O 8.67µl) assembled on ice and incubated at 50°C for 1 hr. PCR product was cloned and sent for sequencing.

2.5.2 Digestion with restriction enzymes

The pGREEN II plasmid (vector map in relevant chapters) was used to constitutively express the MYB8 transcription factor proteins in tobacco and *G. diffusa*. This plasmid contains two copies of the CaMV 35S promoter sequence and a single 35S terminator sequence. The MYB8 coding sequences were amplified from *G. diffusa* cDNA using gene specific primers and the high-fidelity DNA polymerase Phusion (NEB). The amplified products were inserted into pBLUESCRIPT SK- vectors which were used to transform DH5α *E. coli*. The transformed *E. coli* were cultured before the DNA was purified (Section 2.4). The pBLUESCRIPT SK- plasmid and pGREEN II vector were digested using restriction enzymes. 1µg of each vector was mixed with a relevant restriction enzyme/ pair of restriction enzymes (detailed in relevant data chapter). Restriction enzymes were chosen that produced a DNA fragment with the full MYB8 cDNA sequence from the pBLUESCRIPT SK- intact and in the correct orientation for insertion into the pGREEN II vector to enable regulation by the CaMV 35S sequences. The pGREEN II plasmid was cut at one site in between the second 35S promoter sequence and the 35S terminator sequence. The plasmids were digested using the appropriate restriction enzyme/s and buffer (determined using the enzyme manufacturer instructions), incubation was at 37°C for 1 hr 30 mins. The digested products were run on an agarose gel (Section 2.3.6) and the relevant bands, determined by fragment length, were cut out of the gel, and extracted using an Invitrogen extraction kit as detailed in the manufacturer guidelines. The extracted DNA concentration was quantified using a nanodrop (Section 2.3.8). A ligation reaction of 400U of T4 ligase, T4 ligase buffer, and 50ng of vector with a 3:1 insert to vector ratio was used to generate a pGREEN II vector containing a *GdMYB8* insert. Digestion with restriction enzymes was also used to insert promoter fragments into the pHISi vector for use in a yeast one-hybrid assay (detailed in relevant data chapter). Primers to amplify the desired DNA insert from gDNA through PCR were designed with restriction sites for two different enzymes, one restriction site per primer. PCR products then contained the DNA fragment of interest flanked by the restriction sites. The relevant restriction enzymes were used to digest the PCR products and the pHISi vector, digestion and ligation were conducted as described above.

2.5.3 Gateway cloning

Plasmids were constructed for yeast one-hybrid experiments using Gateway cloning (Invitrogen). This method of cloning enables transfer of DNA fragments between plasmids using recombination sequences, while maintaining the reading frame. Primers were designed to amplify a product consisting of the target DNA sequence for vector insertion and two flanking recombination sequences 'attB'. This product was inserted into a pDONR vector through recombination during the 'BP reaction' using BP clonase mix. Once in the pDONR vector, the DNA is flanked with 'attL1' and 'attL2' recombination sites. These pDONR vectors were transformed into DH5α *E. coli,* transformed colonies were grown and plasmid purified (Section 2.4). The Gateway LR recombination reaction was then used to recombine the DNA fragment into the pC-ACT2 vector, which were transferred into *E. coli*, cultured, and the plasmid purified. The reactions were conducted as per the manufacturer's instructions (http://www.igmm.cnrs.fr/wp-content/uploads/2017/11/gateway_pdonr_vectors.pdf).

2.6 Transformation of *Nicotiana tabacum*

(Relevant for Chapters 5 and 6)

2.6.1 *Agrobacterium tumefaciens* strain GV3101

Agrobacterium tumefaciens strain GV3101 was used to transform *Nicotiana tabacum* cv. Samsun to produce transformed plants constitutively expressing MYB8 proteins. The GV3101 strain has a chromosomal rifampicin resistance gene, a gentamicin resistance gene in a disarmed Ti plasmid (pMP90), and tetracycline resistance from a pSOUP plasmid. The Ti plasmid enables tDNA insertion and the pSOUP plasmid promotes replication of the pGREEN plasmid.

2.6.2 Producing electrocompetent *A. tumefaciens*

A. tumefaciens strain GV3101 cells were made electrocompetent. The cells were streaked on an LB agar plate with 25mg/l gentamycin and 50mg/l rifampicin. Following incubation for 2 days at 30°C, a colony was picked and grown in LB media with 10mg/l tetracycline, 25mg/l gentamycin, and 50mg/l rifampicin. The LB culture was grown for two days and then subcultured (1ml culture into 100ml LB media containing 25mg/l gentamycin, and 50mg/l rifampicin) and grown overnight at 30°C. This solution was centrifuged at 3000xg at 4°C for 20 mins to pellet the cells, the supernatant was discarded, and the pellet resuspended in 100ml of cold sterile HEPES. Pelleting and resuspension in HEPES were repeated twice more. The pellet was then resuspended in 8ml of cold sterile 10% (v/v) glycerol, aliquoted, and these aliquots snap frozen in liquid nitrogen and stored in the -80°C freezer. Electrocompetent *A. tumefaciens* cells were transformed with the relevant pGREEN plasmid through electroporation. 1µl of purified 100ng/µl pGREEN plasmid was mixed with 50µl of *A. tumefaciens* cells on ice, prior to electroporation in a Gene Pulser XCell (BIO-RAD). The solution was pipetted into a 2mm GenePulser cuvette (BIO-RAD) and electroporated at 2.4V, 200Ω and a capacitance of 25µF. 1ml of LB was added immediately after electroporation and the cells were incubated at 30°C for 3 - 4 hrs at 180rpm. 50µl of cell culture were then spread onto LB-agar plates containing 25µg/ml gentamicin and 50µg/ml kanamycin. These plates were incubated at 30°C for 48 hrs.

2.6.3 Identifying and culturing transformed *A. tumefaciens*

To identify which *A. tumefaciens* cells were transformed with the plasmid of interest, colony PCRs were conducted on individual colonies. *A. tumefaciens* colonies were picked and mixed with 20µl of 20mM sodium hydroxide, incubated for 10 mins at 37°C and heated for 5 mins to 98°C. 2µl of lysed cells were then used in subsequent colony PCRs as described in Section 2.4. A glycerol stock was created for long term storage of the transformed cells by adding a single transformed colony of *A. tumefaciens* to 5ml of LB containing 25µg/ml gentamicin and 50µg/ml kanamycin, growing the culture overnight at 180rpm and 30°C, adding 50% (v/v) glycerol, shaking the mixture, and then flash freezing in liquid nitrogen. This stock was then stored at -80°C.

A. tumefaciens cells containing the plasmid of interest were prepared for plant transformation as follows: cells from the glycerol stock were streaked onto LB agar plates containing 25µg/ml gentamicin and 50µg/ml kanamycin. The plates were incubated for 2 days at 30°C so that a single colony could be obtained. A liquid culture of *A. tumefaciens* was produced by mixing a single colony with 50ml of LB (with 25µg/ml gentamycin and 50µg/ml kanamycin) in a sterile conical flask and incubating for 24 hrs at 30°C and 180rpm. 1ml of this culture was then added to a fresh conical flask containing the same antibiotic concentrations and grown for 18 hrs. At this point the optical density of the subculture was

measured at 600nm in an Eppendorf Biophotometer AG 22331 – blanked with LB. The subculture was then poured into falcon tubes so that each contained approximately 50ml, these tubes were centrifuged at 10°C and 5000rpm for 5 mins and the supernatant was discarded. The pelleted cells were then resuspended in a volume of sterile half-MS (Appendix 1) that provided an OD_{600} of 0.5. This solution was used to inoculate tobacco leaves during the tobacco transformation procedure described below.

2.6.4 Transforming *Nicotiana tabacum* leaf discs

Healthy, young *N. tabacum* leaves were removed from the plant and sterilised in 10% bleach solution for 15 mins. All subsequent steps were undertaken under sterile conditions. The bleach was removed by rinsing the leaves multiple times with autoclaved DI water. A petri dish was filled with *Agrobacterium* suspension (a cell suspension in half-MS described in Section 2.6.3) and single leaves were submerged in the solution. The edges and midrib of the submerged leaf were removed and, to maximise the surface area of the leaf susceptible to *Agrobacterium* infiltration, leaves were cut into squares (approximately 5 - 7mm across). To remove excess *Agrobacterium*, the leaf squares were transferred to filter paper and then placed adaxial side up on an MS9 plate (Appendix 1, with 0.5mg/ml IAA and 1mg/ml BAP). Once three MS9 plates had been filled with leaf discs, equipment and the *Agrobacterium* solution were changed to minimise the risk of contamination. The MS9 plates containing leaf discs were then incubated for 48 hrs in the dark at room temperature to promote *Agrobacterium* infection of plant cells and consequent integration of a segment of the pGREEN vector into the chromosomal DNA of the host plant. Subsequently, leaf discs were transferred to new MS plates (Appendix 1, with 0.5mg/ml IAA, 1mg/ml BAP, 200µg/ml ampicillin, 100µg/ml kanamycin, and 500µg/ml cefotaxime). The addition of kanamycin to the medium selects against plant cells that do not contain the pGREEN plasmid, enabling mainly transformed cells to regenerate forming callus tissue - from which new transformed plants will grow. All plant tissue was transferred onto fresh MS plates, containing the hormones and antibiotics listed above, every 10 days. Once 1 - 2cm shoots had sprouted from the calli leaf disks produced, these were cut at the base and transferred to 50ml Hamilton jars containing MS9 (with 200µg/ml ampicillin, 100µg/ml kanamycin, and 500µg/ml cefotaxime). After a few weeks these shoots formed roots and, once a sufficient root network had grown, the plants were transferred to plant pots and planted in Levington's M3 bedding compost. Plants were grown in the PGF, kept moist through watering, and infrequently were sprayed with 0.2mg/l Intercept® (Bayer) to control herbivory. Information on genotyping and phenotyping are in the relevant data chapter.

2.7 Analysing gene expression through quantitative real-time PCR

(Relevant for Chapters 4 and 5)

2.7.1 qRT-PCR primer design

The criteria, method of primer design, and tests for primer specificity differed depending on the gene in question and, as such, this information is located in the methods sections of relevant data chapters. Primers were designed using IDT-PrimerQuest (IDT-DNA) to have predicted annealing temperatures of 60°C, GC-content of 40 - 60%, and amplify products between 70 - 140bp in length. Once primers were selected, preliminary qRT-PCRs were conducted, and the resulting melt curve produced from primer products assessed to check for amplification of single products and primer dimerisation. If primer dimers were observed in melt curves, these primers were discarded, and new ones designed.

2.7.2 Calculating qRT-PCR primer pair efficiency

During qRT-PCR, PCR product should double in quantity during the linear phase of DNA amplification. However, efficiency can deviate from 100% and primer pairs showing efficiencies of 90 - 110% were considered suitable for use in subsequent qRT-PCRs. Calculating efficiency also provides a cycle threshold (C_t) range within which products are linearly amplified by primers and enables differences in efficiency to be accounted for in subsequent expression calculations. To calculate primer efficiency a standard curve was produced by carrying out qRT-PCR on plasmids of known concentration. A series of 10-fold serial dilutions of plasmid at 1ng/µl - 10ng/µl in a 100ng/µl yeast tRNA (ThermoFisher) background were produced. For *N. tabacum* qRT-PCR cDNA was used for serial dilutions rather than plasmid DNA (detailed in relevant data chapter). Two technical replicates and a negative control with a template of only yeast tRNA were used in the reaction. The mean C_t value for each plasmid dilution was plotted against the log dilution factor, and a linear line of best fit was calculated. The PCR efficiency was calculated from the slope using the gradient of the best fit line (m): Efficiency = $10^{(-1/m)}$ (Pfaffl 2004). Pairs of primers were selected for use in qRT-PCR for gene expression only if the line of best fit had an $R^2 \geq 0.98$ and efficiency was within the range 1.9 - 2.1.

2.7.3 qRT-PCR analysing candidate gene expression

All qRT-PCRs were completed using Luna Universal qPCR Master Mix (NEB) in a 10µl reaction: 5µl qPCR master mix, 0.25µl forward primer (final concentration 0.25µM), 0.25µl reverse primer (final concentration 0.25µM), 0.3 - 0.6µl cDNA template (of various dilutions, specified in relevant data chapters), sterile DI water up to 10µl. qRT-PCR was conducted in CFX Real-Time PCR machines (CFX384 and CFX Connect) using the following cycle conditions: 95°C for 1 min; 35 cycles of 95°C for 15 secs, 60°C for 30 secs, a plate fluorescence reading; a post-PCR melt curve analysis 60 - 95°C with readings taken at increments of 0.5°C. Data was analysed in the Opticon Monitor software package (BioRad Laboratories, Inc) and the C_t values (i.e. the cycles at which the threshold of fluorescence was crossed) for each well were exported into a Microsoft Excel spreadsheet for statistical analysis.

Reactions were conducted in either 96 or 384 well plates. Three biological replicates of each sample were independently run on separate plates. Each plate contained 3 technical replicate PCR reactions of every sample: cDNA templates from relevant tissues and stages were used to amplify the gene of interest and the reference gene/s along with no template controls. For *G. diffusa* expression analyses the reference gene *Elongation Factor-2* (*GdEF-2*) was selected by Mellers (2016), who assessed the stability of several promising candidates across several floral developmental stages and samples. *GdEF-2* was identified as the most stable using the method outlined in (Chen et al. 2011).

2.7.4 Calculating relative gene expression and data analysis

Data produced from the qRT-PCR reactions were initially checked in the Opticon Monitor software package. Any wells that exhibited abnormal melt-curves and those that had C_t values >1 cycle difference compared to the other technical replicates were discarded. All reactions from a specific primer pair were discarded if there was amplification in the no template control with a C_t value within 6 cycles of the cDNA template reactions. Reactions were also repeated with a higher concentration of template if the C_t values were outside of the tested range of C_t values on the primer efficiency curve. The geometric mean of each set of technical replicates (n=3) was calculated. Using these means, expression of the gene of interest relative to the reference gene/s was calculated through the

efficiency correction method (relative expression = E_{Gene}^{Ct}/ E_{Ref}^{Ct}, where E=primer efficiency, C_t= cycle threshold) (Pfaffl 2004). The geometric mean of relative gene expression across all three biological replicates was then calculated, with standard error and standard deviation calculated using the method outlined in (Ganger et al. 2017). Data were formatted in excel, while graphs and statistical modelling were performed in R (packages:ggplot2 (Wickham 2016), Calculations were completed in excel and graphical representation of the results were completed in R (packages:ggplot2 (Wickham 2016)). Statistical modelling was completed using R packages multcomp (Hothorn et al. 2008) and nlme (Pinheiro et al. 2012).

2.8 Anthocyanin pigment extraction

(Relevant for Chapters 4 and 6)

Pigment extraction was used for approximate quantification of anthocyanin in *G. diffusa* and *N. tabacum* tissues. 20 - 110mg of floral tissue were collected in microcentrifuge tubes, snap frozen in liquid nitrogen, and ground in a tissue lyser (details in Section 2.2.1). Samples were then weighed and 1ml of acidic methanol (1% (v/v) 1M HCL) was added (Day 1). The samples were vortexed and shaken overnight on a platform shaker in the dark. The solution was then spun down, the supernatant was transferred to fresh tubes stored in the dark at -20^{o}C (Day 2). 1ml of acidic methanol was added to the pelleted material, the tubes were vortexed, and left shaking overnight. The supernatant removed on Day 3 was combined with the corresponding supernatant removed on Day 2. Anthocyanin levels were detected using a spectrophotometer measuring absorbance at the wavelengths A530 and A657. This allowed relative anthocyanin content per gram of tissue to be determined. Data analysis was completed in excel and R, detailed in individual data chapters.

Chapter 3. Genetic structure and floral phenotype across the *G. diffusa* Spring range

3.1 Introduction

Establishing the spatial scale of diversification is a key component in understanding species divergence. Reduced gene flow increases the likelihood of genetic differentiation along spatial gradients and it is a necessary prerequisite for speciation (Doebeli and Dieckmann 2003; Slatkin 1973, 1985). This was demonstrated in a meta-analysis of population genetic studies in many taxa, including flowering plants, which found that the probability of speciation occurring within a region tends to decrease as the spatial scale of intraspecific gene flow increases (Kisel and Barraclough 2010). Speciation modes explain the geographical context within which new species form. During allopatric speciation populations are geographically isolated, while speciation in parapatry (between subpopulations) and sympatry (within the same geographic region) occur with initial continued, but reduced, gene flow (Dieckmann and Doebeli 1999; Doebeli and Dieckmann 2003; Heinz et al. 2009). There has been a shift between a perspective centred on these speciation modes to more process-driven research questions (Feder et al. 2013; Fitzpatrick et al. 2009; Ravinet et al. 2017). Many processes can contribute toward speciation including ecological divergence (Peccoud and Simon 2010; Rundle and Nosil 2005), selection through assortative mating (Servedio 2015), runaway sexual selection (Day 2000), evolution following polyploidy (Ramsey and Schemske 1998, 2002), and chromosomal rearrangements (Navarro and Barton 2003; Rieseberg 2001). The methodology we use for defining species may bias our perspective on the relative importance of adaptive and non-adaptive processes for speciation. If species within a biome are predominantly defined based on morphological differences, for example, this may inflate the perceived role of adaptive speciation in promoting diversity within that system. As such, the process of speciation must be investigated at multiple stages along its trajectory from the initiation of genetic differentiation, among populations within existing species, to the evolutionary mechanisms involved in reproductive isolation between populations that ultimately lead to species formation (Ellis et al. 2013; Givnish 2010). While patterns of current and historic gene flow determine the genetic structure of populations, signatures of all mechanisms involved in divergence may not be distinguishable during the late stages of speciation or through sister species comparisons (Ellis et al. 2013; Givnish 2010). Evidently, to understand initial steps driving plant speciation requires characterisation of intraspecific genetic structure, considering genetic divergence between populations, ecotypic variants, and geographic races; doing so in a biodiversity hotspot may also reveal mechanisms underlying high species richness within these ecosystems (Givnish 2010; Musker et al. 2020).

There is a complex interplay between environmental influences and genetic predisposition in determining phenotypic evolution, with both intrinsic and extrinsic factors promoting divergence between populations and eventual speciation (Flatscher et al. 2012). In addition to creating barriers to gene flow, environmental heterogeneity can promote divergence by providing selective environments that differ between populations (Rundle and Nosil 2005; Sobel et al. 2010). This leads to adaptations associated with different ecological niches and habitats, and ecological speciation can occur through reproductive isolation resulting from these adaptations (Schluter 2009). A meta-analysis supporting this provided correlative evidence for the role of ecological adaptation in contributing toward reproductive isolation across diverse taxa, including plants (Funk et al. 2006). Climatic variation, for example, may be an important contributor towards divergence (e.g. Keller and

Seehausen 2012; Wagner et al. 2012) and climatic adaptations can be a key component in speciation (Lowry et al. 2008; Nosil et al. 2005). The spatial scale over which gene flow occurs is also influenced by the specific ecological attributes, life-history traits, and species distribution of each individual taxon – all of which impact population genetic structure (Duminil et al. 2007; Musker et al. 2020). Many of these lineage-specific features have direct or indirect effects on gene flow through pollen or seed dispersal. These features include pollination modes, seed dispersal modes, seed mass, plant size, growth form (i.e. herbaceous or woody, annual or perennial), breeding system (i.e. selfed, mixed or outcrossed), and the extent of the lineage's geographical range (Boucher et al. 2017; Duminil et al. 2007; Givnish 2010; Thomson et al. 2011). Seed and pollen traits that determine pollinator or disperser type can be particularly influential (Dick et al. 2008; Nathan and Muller-landau 2000; Schurr et al. 2005; Vittoz and Engler 2007). Species that self may have reduced effective population size and increased genetic drift, resulting in different alleles becoming fixed in each population (Angeloni et al. 2011; Wright et al. 2008). Conversely, outcrossing might increase within-population gene flow and promote gene flow between populations via pollen export (Duminil et al. 2009; Ellstrand 2014; Gamba and Muchhala 2020). Plant habit can also be an important predictor of population genetic structure; associations have been found between growth form (woody vs. herbaceous) and the level of genetic differentiation between populations, although the mechanism underlying this is not understood (Duminil et al. 2009; Gamba and Muchhala 2020). The impact of lineage-specific traits on dispersal ability may differ between distantly related taxa, for example, variation in seed mass would likely have different impacts on small wind-dispersed dandelion seeds compared to the oak's large animal-dispersed acorns. As such, determining the factors that promote population genetic differentiation requires rigorous understanding of the species, its habit, and the wider environmental context relevant for gene flow.

Namaqualand is a semi-arid desert located within the Succulent Karoo biome of South Africa. This biome has 'a flat to gently undulating topography' (Musker et al. 2020), soils derived predominantly from shale rocks (Bradshaw and Cowling 2014), and some of the highest levels of species richness and endemism among arid plants globally (Cowling et al. 1998). The Succulent Karoo is characterised by low but predictable rainfall that occurs during winter, after which spring annuals bloom in mass flowering events (Cowling et al. 1999; Desmet 2007). Within these communities, there is often a single dominant species with other species interspersed at lower densities (de Waal et al. 2015). Succulent Karoo flora has high numbers of Mesembryanthemaceae, Iridaceae, and Geraniaceae, compared to other arid land floras, and Namaqualand in particular has a high diversity of succulents (Cowling et al. 1999; Cowling and Hilton-Taylor 2001; Desmet and Cowling 1999; Hartmann 1991). Cowling et al. (2009) suggested that the diversity of Cape flora resulted from radiations that were triggered by geomorphic evolution and its interaction with climatic changes in the late Miocene-Pliocene, after which, climatic stability and incremental changes to environmental heterogeneity facilitated high diversification and low extinction rates. While pollinator-driven divergence has been investigated in multiple Cape flora (e.g. Forest et al. 2014; Van Der Niet et al. 2014; Valente et al. 2012), there is a deficiency of research into other mechanisms that contribute toward the initial stages of speciation (Ellis et al. 2013). One notable exception characterised the genetic structure of *Ruschia burtoniae* and *Conophytum calculus*, two co-occurring Aizoaceae species within the Succulent Karoo. The drivers of population divergence were investigated at 'coarse' (10km) and 'fine' (<500m) spatial scales. In *R. burtoniae* strong genetic differentiation at small spatial scales was ascribed to a combination of edaphic specialisation, and low seed and pollen dispersal distances, in a heterogeneous environment.

C. calculus had divergence patterns consistent with genetic isolation by distance. These contrasting within-species genetic patterns indicate that lineage-specific traits are important for determining the spatial scale of gene flow within the system. These results are consistent with previous reports of organism traits (Ihlenfeldt 1994; Klak et al. 2004; Parolin 2001; Parolin 2006) and fine-scale environmental heterogeneity (Ellis and Weis 2006; Ellis et al. 2006) contributing toward population divergence within the Succulent Karoo. Similar studies in additional species within this system are required to disentangle the patterns and processes that are characteristic of Succulent Karoo flora and those that are strongly lineage dependent.

Many of the species within Namaqualand mass flowering communities are Asteraceae. Daisy species in this region form a diverse annual flora, with high incidences of obligate outcrossing, high flower densities, and narrow temporal windows in which to flower (De Waal et al. 2014). Specific life-history traits and pollination modes are associated with Namaqualand daisies and these may be important in determining clade-specific patterns of dispersal; defined as the movement of individuals or propagules with potential effects on spatial gene flow (Ronce 2007). The evolution of daisy species within these communities is significantly influenced by interactions with pollinators (Kemp et al. 2019). A recent study demonstrated that Namaqualand daisies have non-random assembly of complex flower colour patterns, which are functionally associated with specific pollinators. Communities of daisy species within a specific area tend to have one dominant floral colour that varies across the region. The spatial variability of pollinator climates was shown to underlie clustered assembly patterns. Pollinator interaction networks were constructed for two community types (defined by dominant flower colour), one predominantly pollinated by a beefly (*Megapalpus capensis*) and the other by horseflies (*Rhigioglossa* sp.). Community type influenced pollinator interaction strength implying that flower colour traits may be under pollinator selection. Namaqualand daisies were also found to have two distinct syndromes regarding dispersal and selfing ability, in a study investigating annual (wind-dispersed) species. Good dispersal abilities were associated with selfing species, while the propagules of obligate outcrossers had lower dispersal distances. Due to the impact that both breeding system and dispersal ability have on gene flow, evolutionary associations between them are anticipated. Self-compatibility mitigates the risk of pollen limitation for dispersal-prone individuals during the colonisation of new patches or, alternatively, selfing arises due to dispersal consequences. In contrast, outcrossers that are less dispersive are better adapted for persistence in patches (De Waal et al. 2014). These clumped spatial distribution patterns, determined by seed traits and dispersal abilities, can in turn affect biotic interactions by changing pollinator foraging patterns (Hanoteaux et al. 2013) or decreasing the density of heterospecific individuals and thus the amount of heterospecific pollen transfer (Campbell 1986; de Waal et al. 2015). This highlights the dynamic interactions that can occur between processes and factors influencing the spatial scale of gene flow. Pollinator interactions and seed-mediated reproductive assurance are likely to have a prominent role in the evolutionary trajectory of Namaqualand daisy species – with both of these processes directly impacting population genetic structure (Kemp et al. 2019; De Waal et al. 2014).

Gorteria diffusa is a predominantly annual daisy species that grows in large colonies during spring mass flowering events. It is one of multiple Namaqualand daisy species that has parapatric polymorphisms in floral patterning and colouration, others include *Dimorphotheca pinnata* (Thunb.), *Dimorphotheca sinuate*, and *Gazania tenuifolia* (Kemp et al. 2019). Here *G. diffusa* population-level genetic structure was investigated to establish the spatial scale of gene flow within the Spring morphotype, an essential parameter in understanding why *G. diffusa* is so diverse at such small spatial

scales. The extensive intraspecific variation of *G. diffusa* makes it an ideal system for these analyses. The relative phenotypic homogeneity of Spring populations will provide both an interesting comparison to studies in other Succulent Karoo flora, and a counterpoint to investigations of genetic structure across *G. diffusa* morphotypes that have heterogenous floral traits. It will be particularly informative to compare these results with genetic structure analyses conducted between adjacent *G. diffusa* morphotypes where inter-morphotype populations occur over similar spatial scales as those reported here.

Patterns of genetic diversity were quantified using single nucleotide polymorphism (SNP) data obtained from genotyping-by-sequencing (GBS). Focal populations were chosen to encompass both the full geographic area occupied by Spring *G. diffusa* and to provide a range of geographical distances between populations, with the latter enabling representation of gene flow over different spatial scales at high resolution. Population-level genetic differentiation was explored through the analysis of genetic structure using several complementary statistical methods, quantification of between-population genetic differentiation, and investigation into whether isolation by distance patterns were present. Additionally, we conducted a quantitative assessment of floral phenotype in each of the individuals genetically characterised, so that any correlations between genetic and phenotypic variation could be detected. Some variation is evident between the capitula of Spring individuals, but it is currently unknown whether this trait variation is population or region specific. This phenotypic assessment should also provide a strong basis for the subsequent developmental genetic work outlined in this thesis by identifying traits that are fixed within the Spring morphotype. These phenotypic observations may enable initial deductions of likely genetic mechanisms in controlling floral traits. From an evolutionary perspective, establishing with precision which floral traits are both qualitatively and quantitatively consistent within Spring is a prerequisite for evolutionary comparisons between morphotypes considering the mechanisms underlying floral development.

3.2 Methods

3.2.1 Sample collection and preparation

A total of 125 Springbok (Spring) morphotype *G. diffusa* individuals were sampled from the Northern Cape of South Africa in August - September 2018 (Fig 3.2i). All of these individuals were used for subsequent phenotypic analysis ($n_{per\text{-}population}$= 7 - 12) and 75 individuals were used in genotyping-by-sequencing analyses (5 per population). The Spring *G. diffusa* distribution had been previously mapped by Allan Ellis. Sampling locations were chosen to cover most of the distribution of Spring to enable detection of any genetic structure at the within morphotype scale. The experimental design also allowed a range of distances between sampling sites so that dispersal and the role of isolation by distance in the genetic differentiation of the morphotype could be inferred. In each population a minimum of 7 plants were selected for sampling. Within populations plants with adequate numbers of capitula were randomly sampled, ensuring that individuals were at least 5m apart (min = 5.8m, mean = 45.3m, s.d. = 33.8m) to avoid sampling siblings. As multiple plantlets can grow from a single infructescence, each plant was uprooted prior to collection to avoid inadvertently sampling multiple siblings. If multiple plants were present a single plant was selected randomly for use in the analysis. The GPS coordinates of each plant were recorded and two fresh, mature capitula were removed and the stems inserted into a microcentrifuge tube, containing rainwater, for storage before phenotypic analysis. Leaf material was removed from each plant, placed into individual envelopes, and dried using silica gel. Subsequently, flowers were dissected from one sampled capitulum per plant: all spotted ray florets and 3 non-spotted ray florets. These florets were glued to white paper and photographed with an AF-S VR Micro-Nikkor 105mm f/2.8G IF-ED lens mounted on a Nikon D7100 camera with a scale bar included. The distance between the lens and the white paper was standardized.

Dried leaf samples (young leaves of 0.5 – 4cm in length were possible) were collected in paper envelopes containing silica beads and, once dry, additional silica beads were added. The envelopes were then sealed and transported from the field to our laboratory at the University of Cambridge. In Cambridge, the samples were kept in dry conditions and retained in the original envelopes before being used in gDNA extractions. Dried leaf samples were pulverised in a tissue lyser (Qiagen TissueLyser II) and CTAB extractions were conducted as outlined in Section 2.2.2, although slightly modified in that proteinase K was added to the CTAB buffer rather than β-mercaptoethanol. Approximately 20mg of dried leaf tissue was for each extraction. The gDNA purity and concentration were determined using a nanodrop and 2µl of extracted gDNA were run on an agarose gel to ensure that the DNA was not degraded. gDNA extractions of five individuals from each population were selected for genotyping by sequencing analysis. Nanodrop readings indicated that the samples selected had high gDNA content (ng/µl) and high nucleic acid purity (260/280 and 260/230 ratios), while visualisation on an agarose gel showed that gDNA was intact and not fragmented.

Five additional Spring morphotype individuals were included in the DNA extractions and subsequent dataset. These individuals were grown from seeds collected in the field in 2016 (by Allan Ellis). The plants were grown in the Plant Growth Facility at University of Cambridge during 2017 (growth conditions detailed in Section 2.1). No phenotypic data are available from these individuals as they were initially grown and sequenced as part of a different experiment and later co-opted into the current analysis.

3.2.2 Genotyping-by-sequencing and data assembly

Samples were sent to BGI for library preparation and genotyping-by-sequencing analysis. The restriction enzyme ApeKI (G|CWGC) was chosen to digest the genomes as it is a relatively rare cutter. This allowed targeting of a sufficient sequencing depth, considering the relatively large genome size of *G. diffusa* (ca. 2Gb estimated with flow cytometry (Thomas 2009)). Individuals were multiplexed using individual barcodes and the sequencing was performed on a single lane with 150bp paired end reads. Multiplexing was completed for 96 individuals, the 75 individuals from this experiment and an additional 21 for another experiment.

The raw data quality was assessed graphically using diagnostic plots in FastQC (Andrews 2010). Trimgalore (https://github.com/FelixKrueger/TrimGalore) was used to remove adapter content and trim reads to 90bp. The software Stacks v2.41 (Rochette et al. 2019) was used to assemble the genotyping-by-sequencing dataset. As no *Gorteria* genome was available, a *de novo* analysis was performed (the Stacks pipeline is illustrated in Fig 3.1). Loci were clustered through *de novo* analyses using the wrapper program *denovo_map.pl*. This program executes the Stacks pipeline, running each Stacks component. Stacks begins with ustacks, cstacks and sstacks: ustacks identifies sites that are polymorphic within each individual, cstacks connects loci across samples in a catalogue enabling consideration of the metapopulation, and sstacks matches individuals to the catalogued metapopulation data (Catchen et al. 2011). The tsv2bam program stores individual-level data per locus and concludes the locus clustering stage. gstacks uses the set of clustered forward reads (output from ustacks-cstacks-sstacks), and corresponding reverse reads from tsv2bam, to assemble a contig for each locus and create scaffolds. The populations program takes the assembled data from gstacks and applies specified filters to the dataset, outputting VCF files.

Stacks was run on a subset of individuals using different combinations of parameter values, to determine the optimal assembly parameters for the dataset. In order to obtain a representative group of samples, a subset of sampling sites was selected that encompassed the full geographical range of Spring and from each of these sampling sites 4 - 5 individuals were chosen. The range of these parameter values were chosen based on recommendations in Paris et al. 2017. All combinations of the following parameter values (defined in Table 3.1) were tested: m = 3 - 7, M = 1 - 8, n = M - 1, n = M, n = M + 1. The *populations* component of Stacks was run multiple times with an R parameter (R is the percentage of individuals that must possess a particular locus for it to be included in the final dataset) value of either 40, 60, or 80. From the output produced we extracted summary statistics from each Stacks assembly using VCFtools (Danecek et al. 2011), and visualised the data in R using the ggplot2 (Wickham 2016) and tidyverse (Wickham et al. 2019) packages. Using this information, the optimal parameters for running Stacks on the complete dataset were inferred. m = 3 was selected and from the m = 3 datasets the M value that maximised the number of new polymorphic loci before this number plateaued was chosen (M = 5). M = 5 was fixed and the n value that had the highest number of polymorphic loci (n = M + 1 = 6) was selected, following the recommendations in Paris et al. 2017. Stacks was run on the full dataset specifying M5 n6 m3 and R = 80 as parameters. VCFtools (Danecek et al. 2011) was used to filter the data, with the parameters selected listed in Table 3.1.

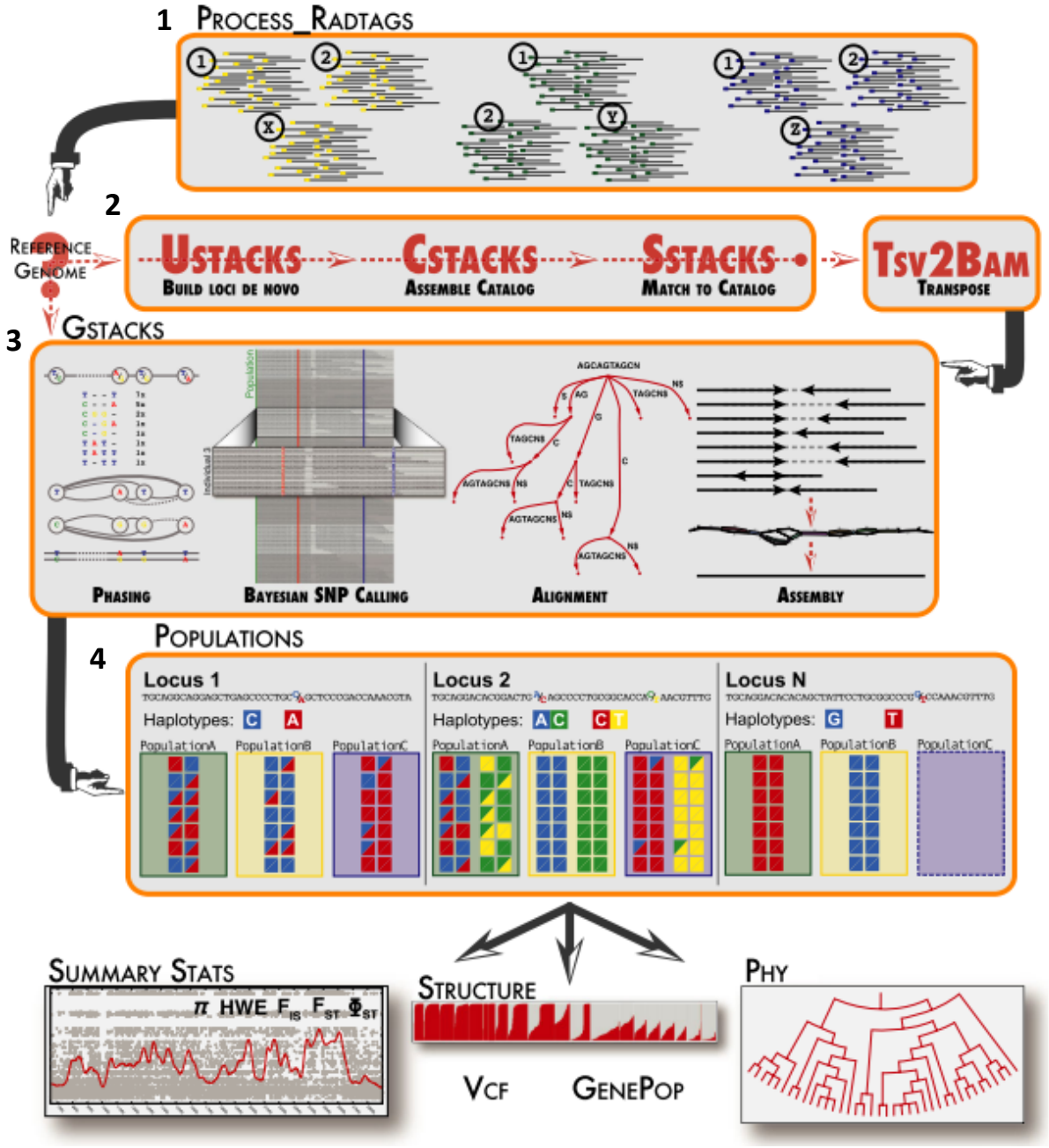


Figure 3.1. Stacks v2 pipeline overview. The four major components of STACKS are indicated by the finger icons: (1) Processing of the sample reads (2) Analyses used for de novo assembly (3) assembling contigs and creating scaffolds (4) Applying a population genetics frame to the data. Figure reprinted with permission by John Wiley and Sons from Rochette et al. (2019).

Parameter	Final Value/s	Definition	Program
m	3	Minimum number of raw reads required to form a stack – a putative allele.	STACKs v2
M	5	Number of mismatches allowed between stacks (putative alleles) to merge them into a putative locus.	STACKs v2
n	6	Number of mismatches allowed between stacks (putative loci) during construction of the catalogue.	STACKs v2
R	80	Percentage of individuals that must possess a particular locus for it to be included in calculation of population-level statistics.	STACKs v2
MIN_DEPTH	8	Only sites with mean depth values (over all included individuals) greater than or equal to the value indicated are included.	VCFtools
MAX_DEPTH	100	Only sites with mean depth values (over all included individuals) less than or equal to the value indicated are included.	VCFtools
max-missing	0.2	Excluding sites where the proportion of missing data is greater than or equal to the value indicated.	VCFtools
hwe	0.05	Assessing sites for Hardy-Weinberg Equilibrium using an exact test (Wigginton et al. 2005) and excluding sites with a p-value less than or equal to the value indicated.	VCFtools
MAF	0, 0.01, 0.05, 0.1	Minor Allele Frequency is the frequency at which the second most common allele occurs in a given population. The value indicates the minimum proportion of the dataset minor alleles have to represent to be retained.	VCFtools

Table 3.1. Parameter values used to filter the genotyping-by-sequencing final dataset.

3.2.3 Diversity estimates

Genetic diversity within populations was assessed by calculating the observed heterozygosity (Nei 1987) using the function basic.stats in the R package hierfstat (Goudet 2005). The observed heterozygosity:

$$Ho = 1 - \sum_{k} \sum_{i} Pkii/np,$$

where *Pkii* represents the proportion of homozygote *i* in sample *k* and *np* the number of samples.

3.2.4 Population genetic structure

Population genetic structure was assessed using 4 types of analysis to assess the pattern of genetic structure and strength of genetic differentiation.

Several datasets were used to explore genetic structure within *G. diffusa* that varied in the specified minimum allele frequency. A single SNP per locus was used to investigate genetic structure to prevent the confounding effects of physical linkage occurring within each locus (Lewontin and Ken-ichi 1960), with the exception of the fineRADstructure analysis described below. This was achieved by thinning the dataset in Stacks so that only one random SNP per locus was retained. As Minor Allele Frequency (MAF) thresholds have been shown to affect population structure analyses (Linck and Battey 2019), we assessed the robustness of our inferences by repeating each analyses on different MAF filtered

datasets: no filtering, 1%, 5% and 10%. The dataset included 75 *G. diffusa* Spring morphotype individuals.

Multivariate analyses

Population structure was explored using principal component analyses performed using the snpgdsPCA function (R package SNPRelate (Zheng et al. 2012)).

Clustering methods

Individual ancestries were investigated using a clustering analysis with a maximum likelihood approach in ADMIXTURE (Alexander et al. 2009). This analysis estimated admixture proportions by determining the proportion of each individual's genome inherited from each of K hypothetical source populations (Gauthier et al. 2020). Files were formatted for ADMIXTURE analysis using the tool set PLINK which converted vcf files into bed files (Purcell et al. 2007; Purcell and Chang 2017). ADMIXTURE was run for K values ranging from 1 - 10, the analysis was completed 20 times for each K value. The model with the greatest maximum likelihood was selected for every K value. The optimal value of K was selected using a 5-fold cross validation procedure. This is an iterative procedure that involves data being partitioned into 5 subsets, admixture estimates are calculated based on 4 of these subsets and used to predict the estimates of the fifth. The prediction error between the expected admixture values and actual values of the fifth subset are then determined. The k value with the lowest cross-validation error is considered to have the best predictive accuracy, regarding the number of subpopulations that make up the total population. Results were visualised using the R package optparse (Davis 2020) to plot cluster assignments for each individual. The script was adapted from one created by Joana Meier (Meier 2019).

fineRADstructure

The program fineRADstructure (Malinsky et al. 2018) was used to infer population structure based on shared ancestry among all individuals. This program is an adaptation of the finestructure program for reduced-representation techniques (Lawson et al. 2012). Co-ancestry estimation is based on whole haplotypes and, therefore, does not require removal of SNPs occurring on the same locus, in contrast to the methods used previously (PCA, admixture). This increases the power of the inference, especially in a system with high numbers of SNPs per locus (evident from transcriptome data) such as *Gorteria diffusa.* To perform this analysis, initially the –radpainter option of the *populations* function of Stacks was used. Second, the *RADpainter* program was run to estimate the co-ancestry matrix between all pairs of individuals. Third, *finestructure* was used to infer relationships between the different populations with 200,000 MCMC iterations (burn-in period of 100,000). Results were then visualized using the *FINERADSTRUCTUREPLOT* R script provided with the fineRADstructure program.

F_{ST} estimates

F_{ST} was calculated in the R package hierfstat (Goudet 2005) using the method and code detailed by Thierry Gosselin (Gosselin 2020). The F_{ST} estimation method used in this package is outlined in Weir and Cockerham, 1984. GPS coordinates of sampled populations were used to create a matrix of geographical distances between every pairwise combination of populations in the package Imap (Wallace 2012). A Mantel test (Mantel 1967) was conducted in R using the package ade4 (Dray and Dufour 2007) to determine whether the relationship between geographical distance and genetic distance was significant among pairs of populations.

3.2.5 Phenotypic analysis

Automated floral trait measurements

Automated trait measurements were completed by Boris Delahaie. In order to obtain precise and quantitative trait measurements for each individual, several dimensions were measured using a standardized picture taken from samples described in Section 3.2.1. Each image was manually cropped to isolate individual elements of the capitulum (individual ray and disc florets). Using the software Fiji (Schindelin et al. 2012), several scripts were designed to perform automatic segmentation of different parts of the flower: whole ray florets and spots. Different variables summarizing the size and shape of these parts were then extracted (Table 3.2). All measurements were standardized to individual flower size to account for individual variation in the size of inflorescences. Ray florets from 182 Spring capitula were processed. Within-individual variation was investigated within 3 populations (H01, M02, R01), with ray florets from 4 capitula per plant analysed from a total of 7 individuals per population. A principal components analysis was conducted on these trait values (Table 3.2) using the function dudi.pca in the R package ade4 (Dray and Dufour 2007). Scatterplots of principal component scores were created using the R package ggfortify (Weir and Goudet 2017).

Trait Measurements
Ratio between the height of the spot and the height of the spotted ray floret
Ratio between the height of the spotted ray floret and the height of the plain ray floret
Aspect ratio of the spot
Aspect ratio of the spotted ray floret
Aspect ratio of the plain ray floret
Circularity of the spot
Circularity of the spotted ray floret
Circularity of the plain ray floret

Table 3.2. The ray floret traits measured and used in a PCA investigating floral trait variation across Spring morphotype individuals.

Quantifying ray floret numbers

From every Spring plant selected for GBS sampling all available capitula were removed. The number of ray florets in total, full spotted ray florets, and partial spotted ray florets (when a complex spot was present but not fully developed) were counted. Two generalised linear model with Poisson distribution and log link function were used to assess whether there was a relationship between spotted ray floret number and total ray floret number. Model 1 had total number of spotted ray florets (full and partial spotted) as the response variable and total number of ray florets as the explanatory variable. In model 2 the total number of full spotted ray florets was the response variable, with number of partial spotted ray florets and total ray floret number as the explanatory variables. To account for the fact that multiple capitula were sampled from the same plant, individual plant was included as a random effect. Statistical analyses were conducted in the R package lme4 (Bates et al. 2015). Checks were conducted to ensure the data met model assumptions, and model simplification was completed using Akaike information criterion (AIC) for stepwise model selection. Graphs were created in the R package fOptions (Wuertz et al. 2017).

3.3 Results

Our genotyping by sequencing dataset consisted of 108.1 million reads distributed relatively evenly between samples of individual plants, with a mean of 1.3 million reads per individual ($n_{individuals}$ = 75, s.d. = 670,000). After initial filtering steps, the dataset contained 295,000 variant sites from 15,500 loci. The mean coverage per individual was 28.5x (s.d. = 6.3x, min = 12.8x, max = 58.2x). MAF threshold has been reported to affect population structure inferences (Linck and Battey 2019), so analyses were run on datasets with MAF 0, 0.01, 0.05, 0.1. As results were fairly consistent between MAF thresholds, all of the results presented in the main text are from analyses of the MAF 0.01 data set and results from analyses of the other MAF data sets are listed in Appendix 3. The filtering criteria outlined in Section 3.2.2 resulted in a final dataset of 75 Spring *G. diffusa* individuals (5 individuals taken from each of 15 populations, with sampling locations illustrated in Fig 3.2i) genotyped with 6,231 SNPs. The retained SNPs had a mean depth of 32.6 (s.d. = 10.4) for MAF 0.01.

3.3.1 Diversity estimates within populations

The within-population observed heterozygosity (H_0) (defined in Section 3.2.3) demonstrated that the level of genetic diversity within each population was relatively consistent (Table 3.3), although SP02 had notably lower heterozygosity. The heterozygosity present was within the range of values reported in the literature for other plant systems (Huang et al. 2017; Lowry et al. 2008; Setsuko et al. 2020).

Population	H_0
H01	0.143
H02	0.145
H03	0.138
H04	0.141
M01	0.148
M02	0.138
M03	0.150
NA10	0.145
NA11	0.136
R01	0.128
R02	0.146
S	0.140
SP01	0.138
SP02	0.115
SP03	0.148

Table 3.3 The observed heterozygosity (H_0) present within each population.

3.3.2 Patterns of population genetic structure across the Spring range

Principal components analysis

To evaluate genetic structure at the level of the individual, a principal component analysis (PCA) was conducted with the genotyping-by-sequencing dataset (Fig 3.2ii). The first three principal components (PCs) accounted for 2.97%, 2.57%, and 2.38% of the variation present, respectively. There was clear genetic structure, with the M cluster (M01, M02, M03) of populations differentiated along PC1 and PC3. SP03 individuals formed a discrete grouping along PC2 and PC3, while individuals from population

H04 showed the greatest genetic differentiation out of all the populations, along PC1. This pattern suggests that genetic structure may be strongly determined by geographical distance, as the patterns of variation are indicative of isolation by distance – with the central cluster of populations lacking clear genetic differentiation and those toward the peripheries of the sampled region (all ≥8km distance from any other sampled site) more differentiated.

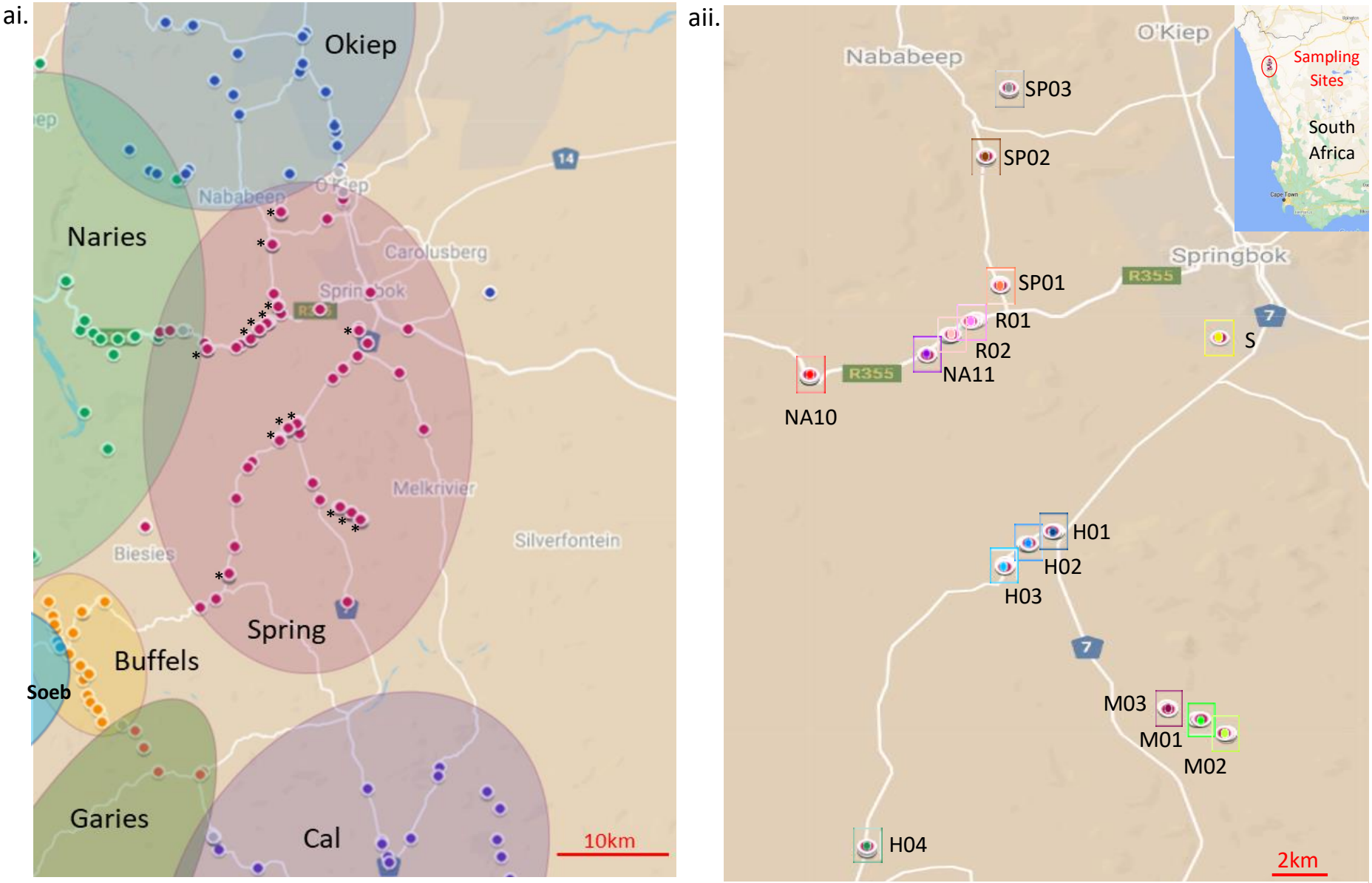
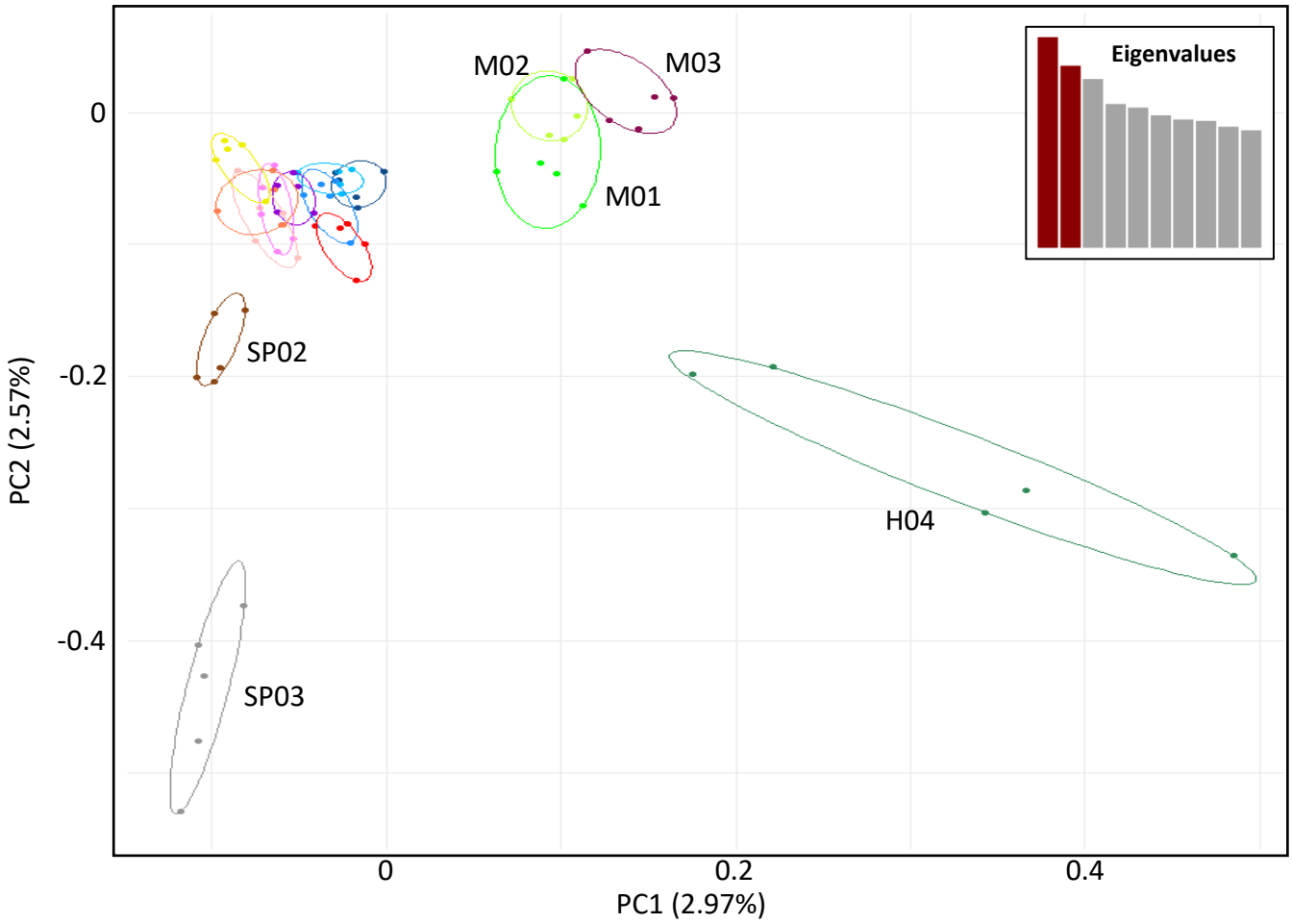
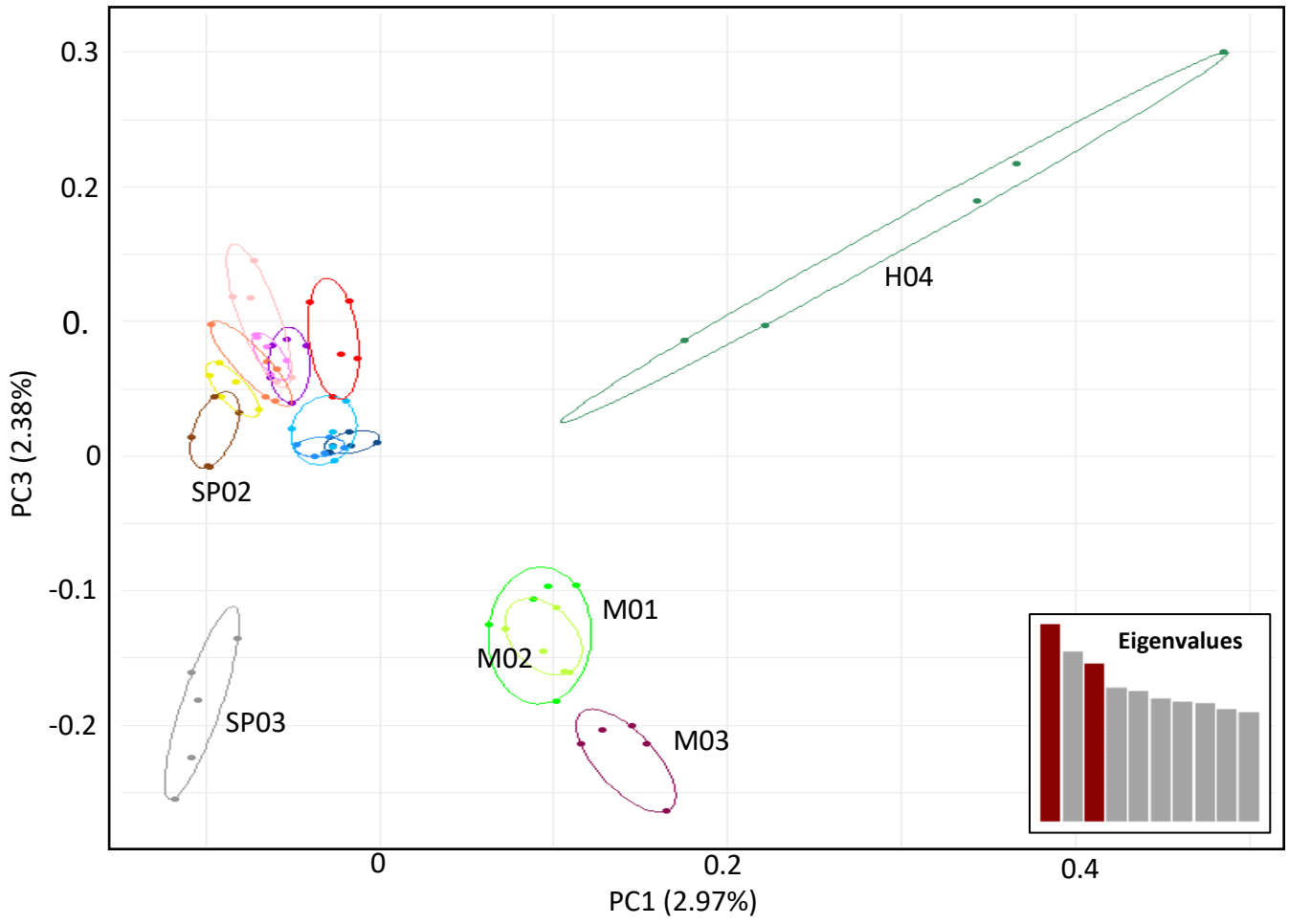


Figure 3.2i. Maps of the Spring morphotype distribution. a) Locations at which Spring and adjacent morphotypes have been found are represented by dots colour-coded by morphotype: Okiep, Naries, Spring, Buffels, Garies, Cal. * populations represent sampling sites. Ellipses provide visualisation of which morphotypes exist in parapatry, rather than precise indications of the range of each morphotype. Grey dots represent locations were hybrid individuals were found. Population locations were provided by Allan Ellis. b) Sampling sites in the Northern Cape of South Africa. Each dot represents a sampled population. Sample site is indicated by a red circle on the map of Western South Africa in the corner. The colour coding of each sampling site matches that in the PC plots of Fig 3.2ii.

b.





H01 H02 H03 H04 M01 M02 M03 NA10 NA11 R01 R02 S SP01 SP02 SP03

Figure 3.2ii. (On previous page) Scatterplots illustrating individual variation in principal component (PC) scores, with values computed from a principal component analysis (PCA). Percentages indicate the amount of variation each PC explains and datapoints are colour coded by population. Eigenvalue graphs demonstrate the percentage of variation accounted for in the first 10 eigenvectors and red bars indicate the eigenvalues represented in the corresponding PC plot. The location of each sampling site is shown in Fig 3.2i.

Clustering analysis

The clustering analysis ADMIXTURE was used to infer individual ancestries. ADMIXTURE uses a statistical model to estimate ancestry proportions from each population, presented as an average over the genome of each individual. The assumption of the model is that the genotype of each individual has ancestry from one or several genetically distinct origins (K). The optimal number of clusters (K) represents the most likely number of ancestral populations and is inferred by the cross-validation errors (CV). Cross-validation errors were lowest at K = 1 (CV = 0.392) (at K = 2, CV = 0.421), and this finding was consistent across different MAF datasets (see Appendix 3). This indicates that across the range of Spring, individuals correspond to one genetic group - so the structure seen is not associated with reproductive isolation. When number of clusters is increased weak genetic structure emerges (Fig 3.3), evident in the tendency of clusters to be divided into geographical location, consistent with the PCA results. The populations in the most Southerly part of the sampling region (H04, M01, M02, M03) cluster at K = 2, and are also the most differentiated in PC1. As the K value is increased the formation of clusters generally tracks the geographic location of each population, for example, at K = 6 there is a Northerly cluster (SP02, SP03), two central clusters (1. NA11, R01, R02, SP01; 2. H01, H02, H03), a South Western cluster (H04) and a South Eastern cluster (M01, M02, M03).

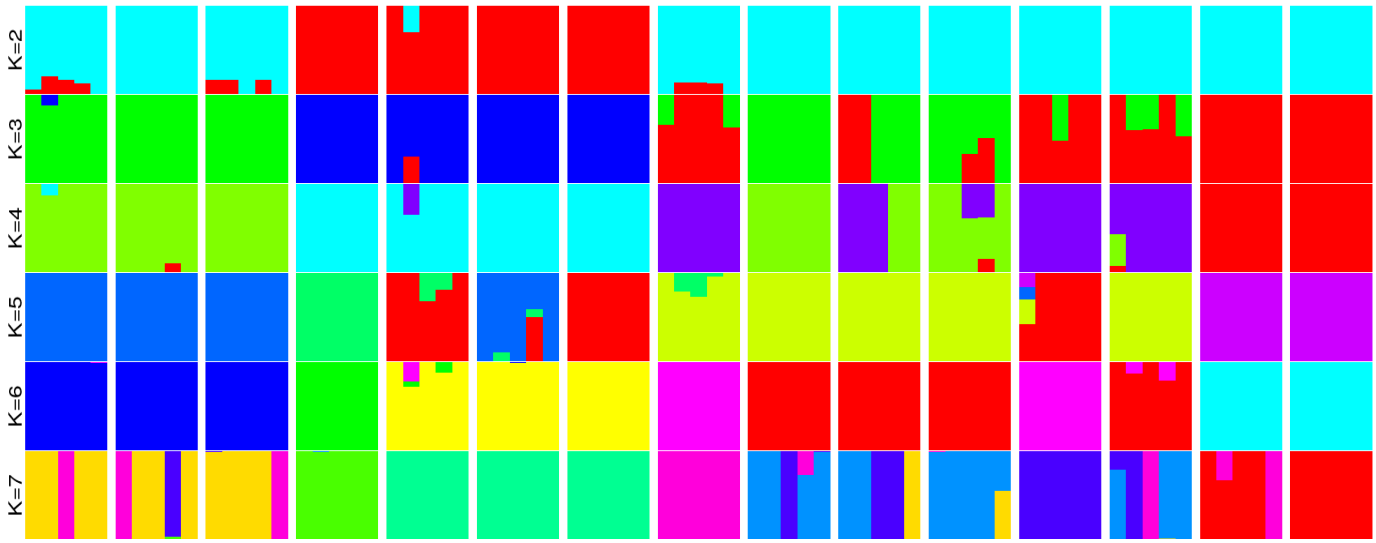


Figure 3.3. An admixture plot demonstrating how the genetic variants in each individual cluster into discrete groupings. The K value indicates the number of clusters specified in the analysis, K = 2 – K = 7 is displayed. Populations are listed along the x axis. Underlined populations are those within the central population cluster identified in the PCA.

Model-based analysis

A model-based Bayesian clustering approach (conducted in fineRADstructure) was used to investigate shared co-ancestry between individuals, enabling population structure inferences. The co-ancestry matrix and cladogram (Fig 3.4) have results consistent with the previous PCA and admixture analyses. As expected, higher levels of shared co-ancestry occur between populations located closer together (i.e. M01, M02, M03; H01, H02, H03; SP02, SP03; R01, R02, NA11, SP01) compared with those more geographically isolated. The central population cluster identified in the PCA is also visible here; readily

distinguished from populations on the peripheries of the Spring morphotype range (H04, M01, M02, M03, SP03) by elevated shared co-ancestry levels within this central cluster. The cladogram also groups this central cluster of populations.

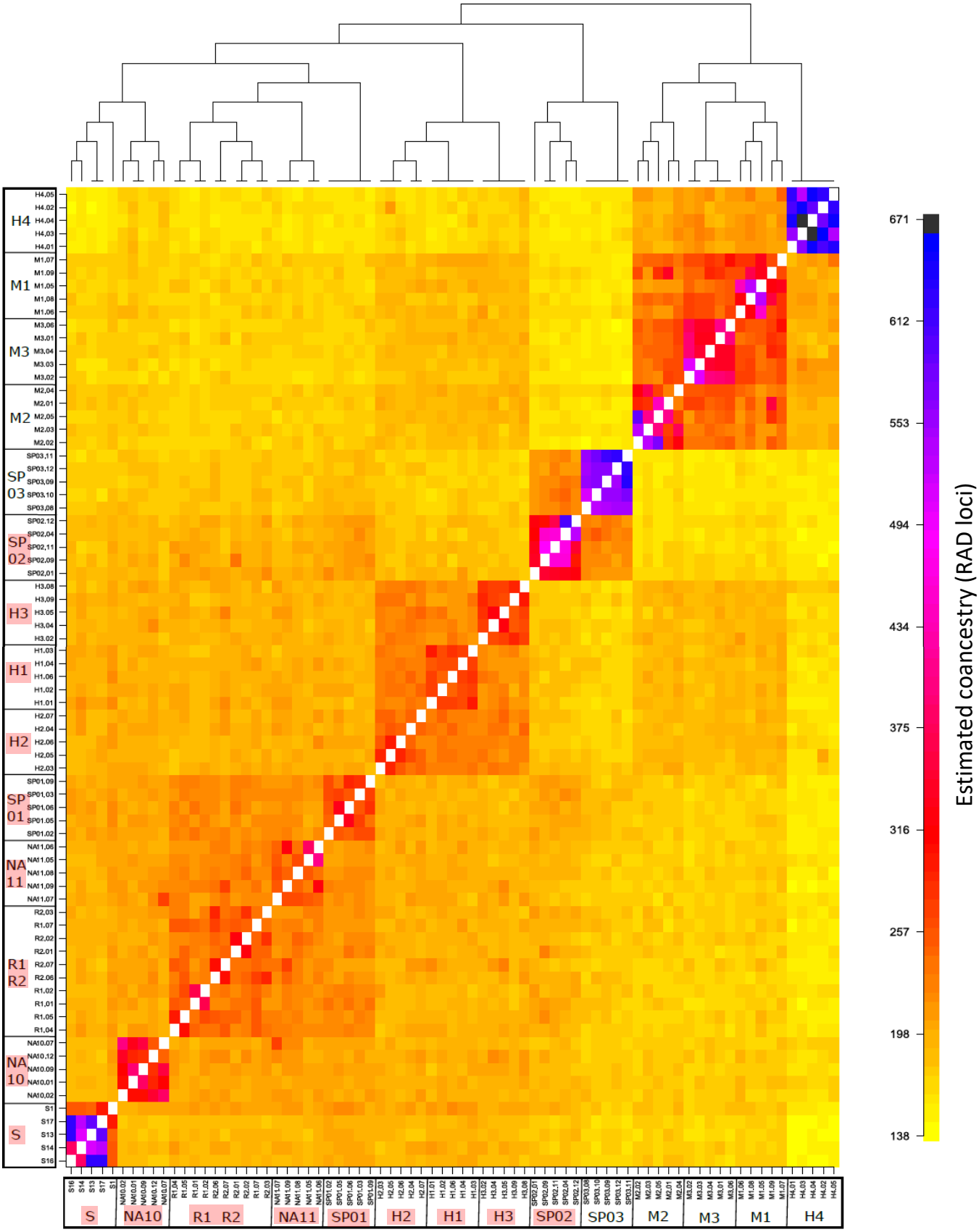
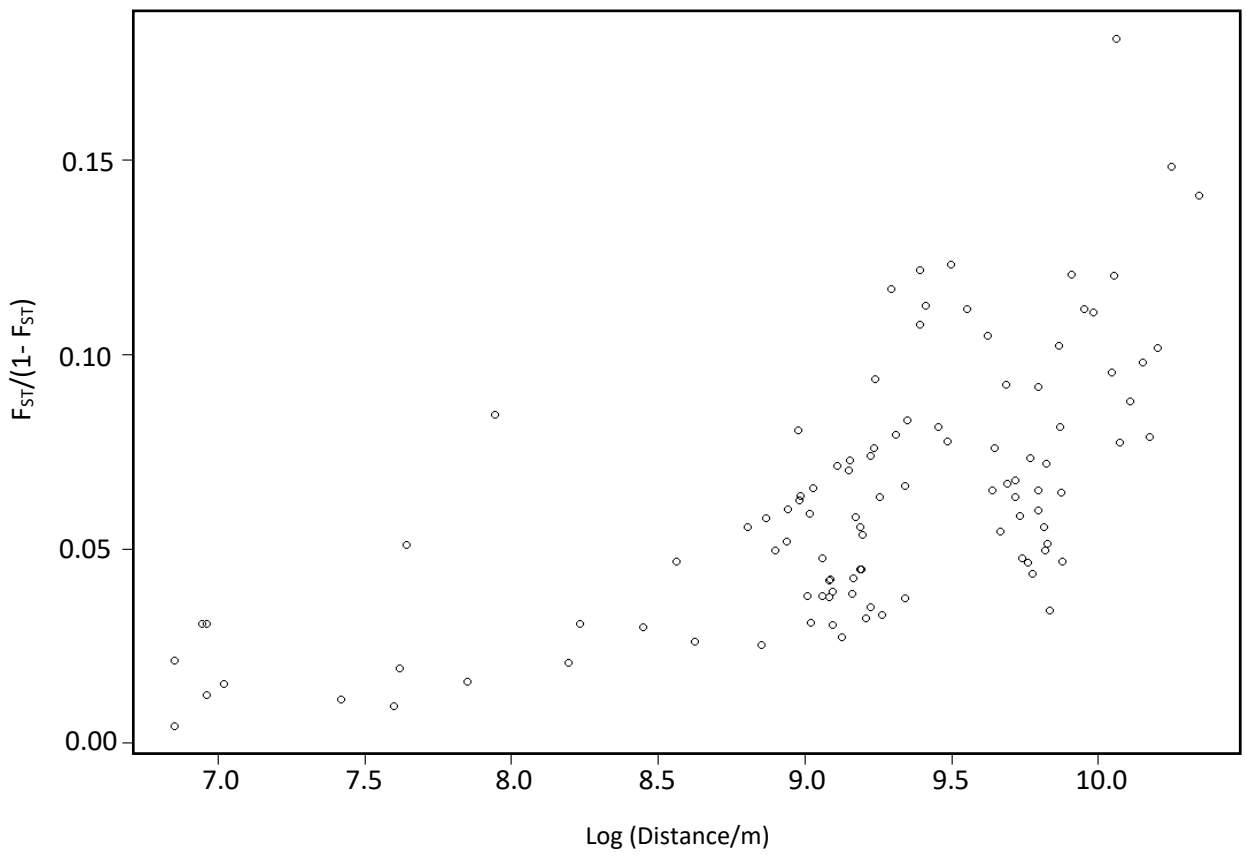


Figure 3.4. Clustered fineRADstructure coancestry matrix. Individuals are listed along with the population from which they derive. The colour of each square represents the results of a pairwise comparison of estimated coancestry between two individuals based on similarity between RAD loci. The relative coancestry is illustrated, with high levels indicated by blue colouration and lower levels indicated by yellow. The phylogeny is illustrative of relationships between populations but does not represent true population history. Populations highlighted in red are those within the central population cluster identified in the PCA.

3.3.3 Genetic differentiation

Genome-wide differentiation was measured using the F_{ST} estimation outlined in Weir and Cockerham, 1984. F_{ST} can be "regarded as the correlation due to common ancestry between random gametes from the same subpopulation relative to random gametes from the total population" (Wang 2012). Overall genetic differentiation averaged across all populations gave an F_{ST} of 0.058 (CI: 2.5% = 0.055, 97.5% = 0.061) (Weir and Cockerham 1984). Between-population pairwise F_{ST} was positively correlated with geographical distance (Mantel's r = 0.36, p<0.0001, 9999 permutations), indicating an isolation by distance pattern of divergence (Fig 3.5a). Some pairs of populations had relatively high genetic differentiation at small spatial scales, relative to the general trend, including SP02-SP03 (F_{ST} = 0.078, Distance = 2.82km), M02-M03 (F_{ST} = 0.048, Distance = 2.09km), and H04 with several populations (H04 and: H01 F_{ST} = 0.10, Distance = 14.08km; H02 F_{ST} = 0.11, Distance = 13.31km; H03 F_{ST} = 0.10, Distance = 12.24km; M03 F_{ST} = 0.10, Distance = 10.87km). M01 - SP01 showed particularly low genetic differentiation given how far apart they were geographically (F_{ST} = 0.033, Distance = 18.63km). The vast majority of pairwise F_{ST} values were significantly different from zero (as confidence intervals did not overlap with zero) (Fig 3.5b). Beyond a distance of 2km significant genetic differentiation was always detected between populations. 5 population pairs were not significantly genetically differentiated: (1) H02 - H03 1.12km, (2) R01 - NA11 1.99km, (3) R02 - NA11 1.06km, (4) R01 - R02 0.95km, and (5) R01 - SP01 1.67km. There were 3 population pairs in closer proximity to one another than (2) - (5), that were significantly genetically differentiated: H01 - H02 (F_{ST} = 0.021, Distance = 0.95km), M01 - M03 (F_{ST} = 0.030, Distance = 1.04km), and M01 - M02 (F_{ST} = 0.030, Distance = 1.06km). The highest levels of genetic differentiation were between H04 and S, H04 and SP02, and H04 and SP03 (all have F_{ST}>0.12). Again, this is consistent with isolation by distance for SP02 and SP03, as these populations are the most Northerly and H04 is the furthest South. H04 is 23.4km from S and, as S individuals were collected during a different year from the other samples, this high level of genetic variation may also be influenced a temporal component affecting genetic differentiation. Overall, these results show that there is strong genetic differentiation over small spatial scales within the Spring morphotype of *G. diffusa*.

a.



b.

	H01	H02	H03	H04	M01	M02	M03	NA10	NA11	R01	R02	S	SP01	SP02	SP03
H01		0.021	0.019	0.100	0.040	0.053	0.060	0.043	0.037	0.030	0.037	0.066	0.034	0.061	0.084
H02	0.007 0.038		**0.015**	0.110	0.040	0.051	0.059	0.041	0.030	0.027	0.036	0.070	0.032	0.052	0.067
H03	0.007 0.028	**0.000** **0.027**		0.101	0.045	0.055	0.057	0.043	0.038	0.031	0.037	0.062	0.036	0.060	0.075
H04	0.083 0.115	0.095 0.125	0.086 0.113		0.077	0.097	0.105	0.093	0.108	0.100	0.100	0.153	0.107	0.129	0.123
M01	0.030 0.054	0.026 0.057	0.037 0.056	0.063 0.090		0.030	0.030	0.047	0.045	0.042	0.044	0.071	0.033	0.072	0.073
M02	0.042 0.066	0.038 0.066	0.045 0.066	0.085 0.110	0.020 0.039		0.048	0.061	0.057	0.049	0.053	0.085	0.045	0.081	0.092
M03	0.049 0.074	0.046 0.073	0.046 0.069	0.092 0.117	0.021 0.041	0.037 0.061		0.068	0.062	0.055	0.063	0.095	0.061	0.087	0.089
NA10	0.034 0.058	0.029 0.057	0.033 0.053	0.081 0.105	0.037 0.058	0.050 0.072	0.058 0.077		0.030	0.025	0.029	0.075	0.025	0.060	0.072
NA11	0.027 0.047	0.014 0.043	0.025 0.046	0.093 0.119	0.034 0.055	0.045 0.069	0.048 0.070	0.019 0.038		**0.009**	**0.012**	0.067	0.020	0.056	0.074
R01	0.014 0.043	0.009 0.041	0.021 0.043	0.084 0.114	0.032 0.051	0.038 0.060	0.043 0.064	0.013 0.035	**-0.001** **0.018**		**0.004**	0.049	**0.011**	0.053	0.068
R02	0.027 0.047	0.024 0.053	0.027 0.045	0.089 0.113	0.036 0.054	0.043 0.062	0.052 0.076	0.021 0.038	**0.003** **0.021**	**-0.003** **0.014**		0.062	0.016	0.047	0.069
S	0.053 0.078	0.054 0.086	0.051 0.071	0.140 0.166	0.059 0.080	0.073 0.097	0.083 0.107	0.067 0.085	0.056 0.077	0.040 0.059	0.051 0.074		0.055	0.086	0.108
SP01	0.021 0.046	0.017 0.045	0.025 0.043	0.093 0.122	0.025 0.043	0.037 0.054	0.052 0.073	0.017 0.033	0.009 0.031	**0.004** **0.021**	0.008 0.023	0.042 0.066		0.045	0.075
SP02	0.051 0.076	0.035 0.065	0.049 0.072	0.113 0.139	0.059 0.085	0.070 0.092	0.074 0.099	0.048 0.073	0.043 0.067	0.040 0.065	0.039 0.056	0.070 0.098	0.035 0.053		0.078
SP03	0.072 0.100	0.052 0.083	0.063 0.090	0.108 0.137	0.062 0.082	0.077 0.104	0.079 0.098	0.059 0.086	0.062 0.086	0.054 0.083	0.058 0.082	0.096 0.119	0.060 0.089	0.065 0.093	

Mean F_{ST} value

≥0.11	≥0.09	≥0.07	≥0.05	≥0.03	≥0.02	<0.02

Figure 3.5. Pairwise F_{ST} values demonstrating genetic differentiation between populations. a) Pairwise F_{ST} plotted against geographical distance between the 2 populations. b) Matrix of pairwise F_{ST}, mean F_{ST} values are presented in upper right section and confidence intervals (2.5%, 97.5%) in the lower left section. Values are colour-coded according to mean F_{ST} for ease of comparison, for example, light yellow boxes have 0.03 > F_{ST} ≥0.02. Values underlined and in bold have a lower confidence interval value that is <0.005 and genetic differentiation between these populations, as measured by pairwise F_{ST}, is considered non-significant.

3.4 Variation in phenotype

Floral trait measurements were compared between populations to determine whether there was a relationship between genetic differentiation and floral phenotypic differences. To ensure that single trait measurements per individual were representative of variation within a plant, multiple measurements were taken from individuals within a subset of populations. To determine the suitability of specific traits for comparisons between localities, qualitative trait assessments were conducted within an individual.

Ray floret trait measurements from Spring capitula were extracted from photographs taken in the field. The measurements included size and shape variables of petal spots and ray florets (detailed in Table 3.2). These variables were combined in a principal component analysis to investigate whether there was phenotypic differentiation over the morphotype range. PC1 explained 45% of the variance and PC2 19% (PC loadings are given in Table 3.4). When Spring flowers were grouped according to locality, no signal of floral phenotypic differentiation was found between populations (Fig 3.6, Supp. Table 2). Phenotypic variation between individuals within a population was investigated in 3 populations (Fig 3.7), with multiple capitula sampled per plant. As expected, capitula within individuals were more similar than those between. There was often overlap between the PC scores of floral traits from different individuals.

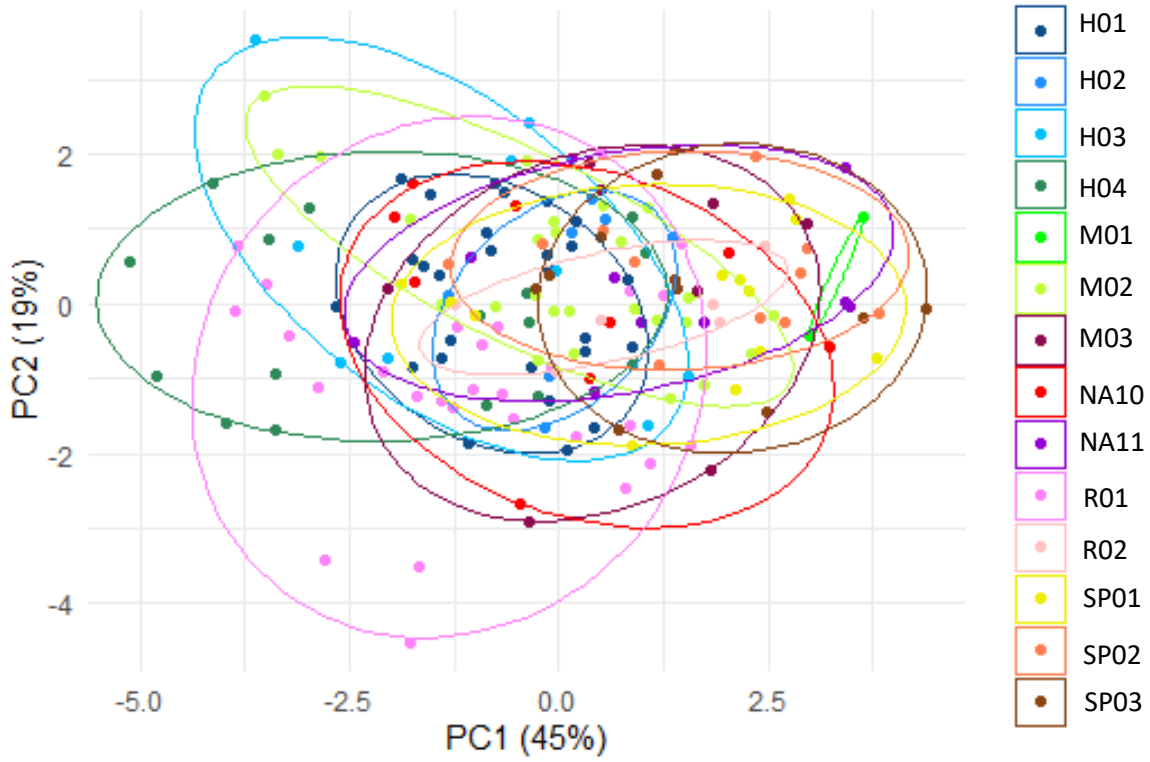


Figure 3.6. Scatterplot illustrating individual variation in ray floret measurements. Principal component (PC) scores are along each axis, with values computed from a principal component analysis (PCA). Percentages indicate the amount of variation each PC explains and datapoints are colour coded by population.

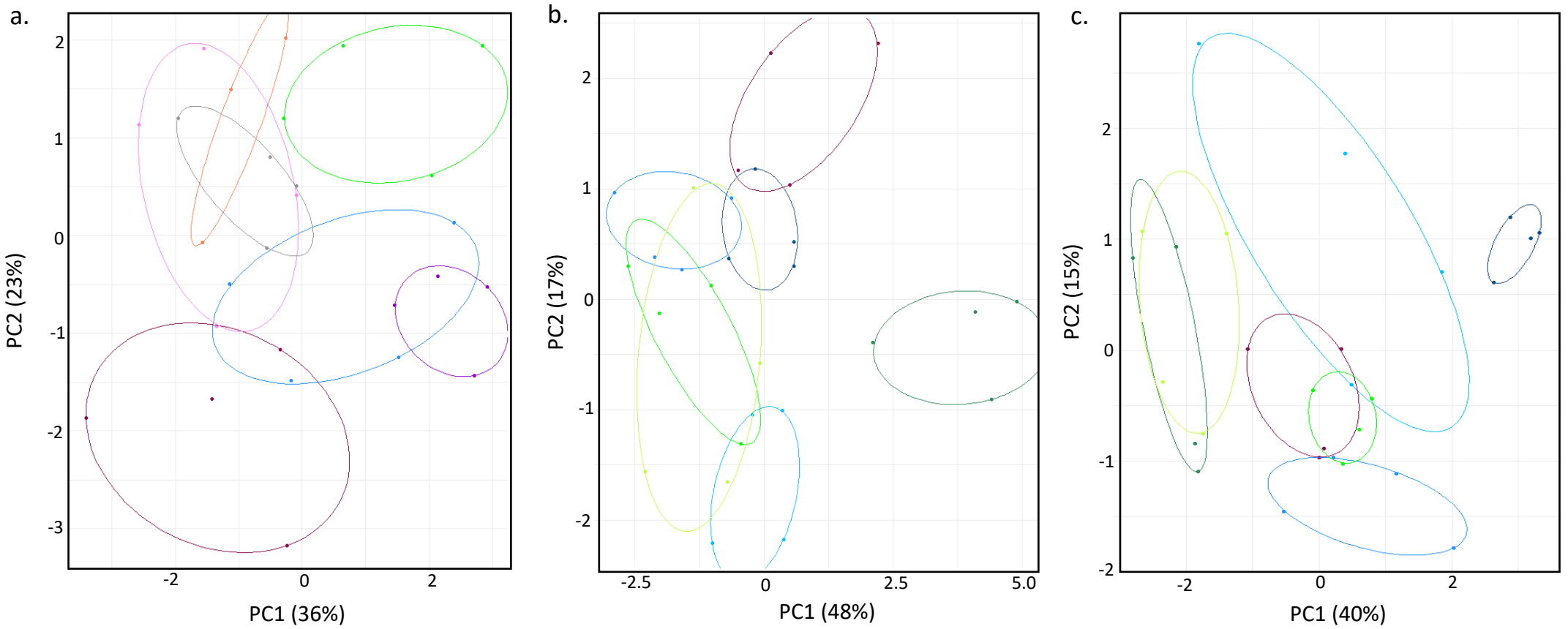


Figure 3.7. Scatterplot illustrating variation in ray floret measurements between capitula. In population a) H01 b) M02 c) R01. Principal component (PC) scores are along each axis, with values computed from a principal component analysis (PCA). Percentages indicate the amount of variation each PC explains and datapoints are colour coded by individual.

Trait Number	PC1	PC2
1	0.195	0.698
2	-0.168	-0.271
3	-0.488	0.0869
4	-0.429	-0.217
5	0.446	-0.162
6	0.430	0.087
7	-0.311	0.586
8	0.175	-0.086

Table 3.4. The relative loadings of individual variables on each principal component (PC1 and PC2) for the analysis presented in Fig 3.6. Traits are as follows: (1) Ratio between the height of the spot and the spotted ray floret, (2) Ratio between the height of the spotted ray and the plain ray floret, (3) Aspect ratio of the spot, (4) Aspect ratio of the spotted ray floret, (5) Aspect ratio of the plain ray floret (6) Circularity of the spot, (7) Circularity of the spotted ray floret, (8) Circularity of the plain ray floret.

Quantifying ray floret numbers in Spring

The number of spotted ray florets within a capitulum does vary between capitula within a single Spring plant (Fig 3.8a), as such this is not a suitable trait for between - locality phenotypic comparisons based on static floral traits. In some individuals, all plain (non-spotted) ray floret petals have simple basal spots ('marks') – comprising dark pigmentation but no epidermal elaborations (Fig 3.8b). A thorough quantification of ray floret numbers and presence or absence of marks has not previously been conducted.

These traits were quantified in individuals across multiple populations and all capitula available in each individual were included ($n_{populations} = 14$, $n_{individuals} = 122$, $n_{capitula} = 416$). The presence or absence of marks at the base of plain ray floret petals was always consistent within an individual, either all capitula had petals with marks or the trait was completely absent across all capitula. Overall, 73% of individuals had the mark phenotype present within ray floret petals. Sample sizes within populations were too small to assess within-population patterns. This would be an interesting focal trait for phenotypic comparisons between localities, given that the genetics underlying spot pigmentation are currently being characterised. There was a mixture of individuals with and without marks found in 11 populations and in the remaining 3 populations all individuals had marks. As such, no population included consisted solely of individuals lacking the mark phenotype.

The number of spotted ray florets was significantly higher in capitula with a greater total number of ray florets, both when only fully developed spots were taken into account ($z = 2.9$, d.f. = 413, $p < 0.005$) and when all spotted ray florets (full and partial spots) were considered ($z = 3.1$, d.f. = 413, $p < 0.002$) (Fig 3.8c). 'Partial spots' here refers to the phenotype when complex spots begin to form, but do not fully develop (Fig 3.8b). When there was a greater number of fully developed spots within a capitula, there were fewer partial spots ($z = -2.3$, d.f. = 413, $p < 0.02$).

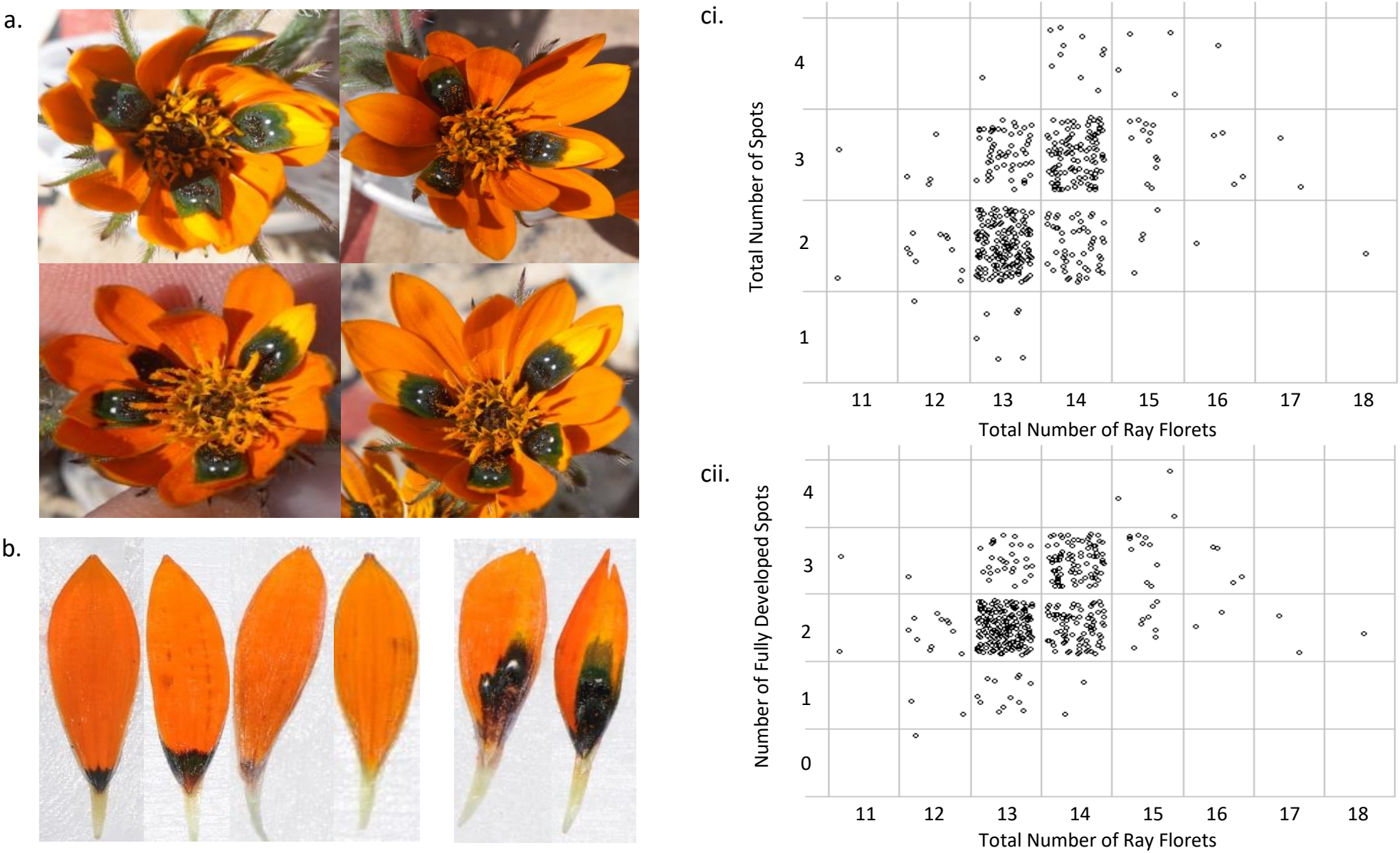


Figure 3.8. Quantifying spot and ray floret number. a) The capitula from one Spring morphotype individual (M02.01), note that spot number varies between capitula. b) Examples of variation in plain ray floret phenotypes (left-hand panel), simple basal spots (marks) are present in 3 out of 4 of the ray florets shown – each taken from a different individual plant. Example of partial spots (right-hand panel) where complex spot development is initiated but does not reach completion. c) Plots demonstrating the relationship between the total number of ray florets in a capitulum and i) total spotted ray floret number (spot and partial spots) ii) total number of spotted ray florets with fully developed spots. Each data point is a single capitulum. Photographs were taken by Boris Delahaie.

3.4 Discussion

Our results show that there is genetic structure within the Spring morphotype of *G. diffusa* across its native range. The patterns of genetic variation were consistent with isolation by distance, whereby individuals that are closer together geographically tend to be more genetically similar due to limited dispersal (Slatkin 1993; Wright 1943). There was strong genetic differentiation at the population level (global F_{ST} = 0.058) considering the relatively small spatial scale, with distances between populations ranging from 0.95 - 31km. Floral phenotypic characterisation demonstrated that variation in floral traits between individuals was not clustered in specific populations or regions. Both total ray floret number and spotted ray floret number varied between capitula within an individual, demonstrating that these traits were not suitable for assessing between population phenotypic variation. The positive correlation between the number of ray florets and the number of petal spots within a capitula has implications for models of *G. diffusa* capitulum development.

The Spring morphotype of *G. diffusa* has clear genetic structure across its native range. The genetic variation follows a predictable geographic trend, evident in the positive correlation between pairwise F_{ST} and between-population geographic distance. Despite obvious genetic structure, the clustering algorithm admixture demonstrated that the Spring morphotype does not comprise distinct genetic groupings. This provides further support that the genetic structure of Spring results from an isolation by distance pattern, rather than differentiation resulting from a barrier to gene flow. The most Southerly population (H04) was consistently more genetically differentiated than other populations across multiple analyses and, to a lesser extent, the most Northerly population also had relatively higher genetic differentiation. Heterozygosity within H04 and SP03 was consistent with that found in other populations and so this higher level of genetic differentiation cannot be accounted for by unusual within-population genetic diversity patterns. It is possible that this differentiation is simply due to these populations being further geographically from the majority of sampling sites. However, the two most Northerly populations had 2.82km between them and a pairwise F_{ST} of 0.078. Given that one of these populations is approximately 5km from the Okiep morphotype range, this high differentiation may have, in part, resulted from introgression of Okiep morphotype genes into one Spring population. Our experimental design incorporated several sets of populations with equivalent distances between them (Fig 3.2i), which provided multiple comparisons for investigation of the spatial scale of gene flow within the Spring range. Of our six populations pairs <1.2km apart, half exhibited levels of genetic differentiation significantly different from zero. This indicates that genetic structure was present at even finer spatial scales than our study design incorporated.

The isolation by distance pattern found here was consistent with our expectations based on the consistency of Spring floral phenotypes, and observations of landscape characteristics across the morphotype range. The Spring morphotype inhabits the Kamiesberg region of Namaqualand, an area with rolling hills and in-filled valleys between them (Desmet 2007). There was no obvious environmental heterogeneity present within different regions, for example, topographical differences, climatic variation, or substantial differences in community structure. Similarly, no potential barriers to gene flow like rivers, mountains, or different drainage system, were identified. Of course the geographically homogenous floral phenotypes of Spring do not preclude the existence of cryptic diversity (Britton et al. 2014; Krejčíková et al. 2013) and there could be undetected environmental gradients, for example, fine scale edaphic heterogeneity (Ellis and Weis 2006; Ellis et al. 2006).

Isolation by distance is a mechanism found to contribute to genetic structure in other Succulent Karoo flora including *Conophytum calculus* (Musker et al. 2020) and *Protea repens* L (Prunier et al. 2017). *G. diffusa* is an obligate outcrosser, a trait associated with lower dispersal abilities in a study of Namaqualand daisy species (De Waal et al. 2014). The dispersal ability of *G. diffusa* diaspores have not been investigated, but inferences can be made from diaspore phenotypes and field observations. After floral anthesis, *Gorteria* receptacles lignify and enclose the fruit, these infructescences then drop off and achenes germinate from within the infructescence (Duncan and Ellis 2011; Karis et al. 2009). These infructescences seem to disperse primarily along the ground, moved by the wind or rolling down slopes and the lignified involucre of the infructescence also become entangled in foliage. As such, it is likely that seed dispersal could be inhibited even by small obstructions. These effects are evident in the clumped distributions of *G. diffusa* plants and high occurrences of multiple plants growing at the base of shrubs and boulders, and within rock crevices. The high numbers of infructescences at the base of bushes suggests that these plant distribution patterns are not due to grazers feeding mainly on plants that grow in the open, although this could also be a contributing factor. Limited seed dispersal could, therefore, be a major factor contributing toward strong isolation by distance patterns and high genetic differentiation in *G. diffusa*.

The dominant pollinator of *G. diffusa* is the bee-fly *Megapalpus capensis* and pollen movement may contribute toward connectivity between populations (Ellis and Johnson 2009). It is thought that these flies do not transfer pollen over long distances but tend to stay localised to specific areas. *M. capensis* feed on Spring capitula and this morphotype has petal spots which are sexually deceptive to males, inducing inspection behaviour and pseudocopulatory responses (Ellis and Johnson 2010; Johnson and Midgley 1997). While all of these behaviours result in pollen transfer, sexual deceptive responses may enhance pollen export. An experiment supported this, showing higher rates of fluorescent powder (a pollen analogue) transfer occurred between Spring capitula when the pollinator was exhibiting mate-seeking, rather than feeding, behaviour (Ellis and Johnson 2010). The contribution of pollen dispersal by *M. capensis* to gene flow may, therefore, vary between morphotypes depending on whether or not floral phenotype induces sexually deceptive responses in male flies. Flowering phenology could also influence pollen dispersal patterns. *G. diffusa* germinates in winter, during the predictable winter rainfall characteristic of the Succulent Karoo, and flowers in Spring (Cowling et al. 1999; Desmet 2007). The rain showers triggering gemination do not necessarily occur simultaneously across the Spring morphotype range and, during sampling, the growth stage of *G. diffusa* plants differed between sites. These temporal differences in flowering time could reduce gene flow between geographical areas; as flowers may not be receptive to pollen or producing pollen within the same timeframe. The consequent reduction in cross-pollination between sites, caused by incomplete overlap in flowering times, could enhance genetic differentiation.

Overall, population-level genetic differentiation was strong (global F_{ST} = 0.058), given the relatively small sampling area (0.95 - 31km). *G. diffusa* is comprised of multiple floral morphotypes existing in parapatry. Due to this high intraspecific diversity and the steepness of clines between floral forms (Ellis and Johnson 2010), it was anticipated that *G. diffusa* would exhibit high levels of genetic differentiation over relatively small spatial scales. Our findings were consistent with this hypothesis and the within-morphotype spatial scale of gene flow was of particular interest. Studies investigating population genetic structure in the Cape Floristic Region (adjacent to the Succulent Karoo), reported similar levels of genetic differentiation but across much broader geographical areas. *Protea repens* (Proteaceae), for example, had a global F_{ST} = 0.063 with some sampling sites >800km apart (Prunier

et al. 2017). The sampling distribution for *Restio capensis* (Restionaceae) was >500km and global F_{ST} = 0.030 (Lexer et al. 2014), while the *Seriphium plumosum* complex (Asteraceae) also had a large span of sampling sites and F_{ST} values 0.004 - 0.061 (Shaik 2019). The pattern of high differentiation found in Spring *G. diffusa* was consistent with two additional studies conducted within the Succulent Karoo considering genetic structure of Aizocaeae species at relatively small spatial scales. *Argyroderma pearsonii* had sampling locations <10km apart and a global F_{ST} = 0.07 (Ellis et al. 2007), while an equivalent global F_{ST} = 0.068 in *Ruschia burtoniae*, from sites spanning 17 - 42km, was considered strong population-level differentiation by the authors (Musker et al. 2020). In the latter, *Conophytum calculus* was investigated in parallel over the same geographic range and exhibited weak genetic differentiation (F_{ST} = 0.009). Similarly, in *Protea repens* (Prunier et al. 2017) and *Restio capensis* (Lexer et al. 2014) adaptive processes contributed toward genetic differentiation, but environmental differences seemed to be driving speciation exclusively in *R. capensis*. The differential responses of these species is thought to derive from taxon-specific features, such as sensitivity to specific environmental gradients and contrasting pollination systems impacting population connectivity (Musker et al. 2020; Prunier et al. 2017). These lineage-specific idiosyncrasies are influential in determining the spatial scale of gene flow, and caution against making generalisations on the basis of our findings. However, high genetic differentiation within Spring *G. diffusa* is consistent with the hypothesis proposed by Musker et al. (2020) that Succulent Karoo plants may have finer-scale differentiation then plants from the neighbouring Fynbos biome. To address this hypothesis would require further studies in many additional species.

Evidently, strong isolation by distance patterns are a major factor influencing genetic structure in Spring *G. diffusa*. Genetic differentiation is present at fine spatial scales, in some cases between populations only 1km apart. No clear genetic clustering was observed within Spring, suggesting genetic connectivity between populations is limited primarily by geographic distance resulting from an intrinsic limit to gene flow rather than extrinsic barriers that would create a non-clinal pattern of structure. We hypothesise that these patterns result mainly from low dispersal abilities of *G. diffusa* infructescences. A more comprehensive understanding of the relative influence of seed and pollen dispersal on genetic structure would provide greater clarification. Characterising Spring *G. diffusa* seed and pollen dispersal distances, alongside analyses into the relative contribution of maternal and biparentally inherited loci to gene flow, could resolve these questions. Further analyses investigating the directionality of gene flow between sites would determine if genetic patterns result from ongoing gene flow. The alternative is that populations were initially isolated and are becoming more homogenous through recent range expansion (Peter and Slatkin 2013; Turelli et al. 2001).The spatial restriction of gene flow can promote divergence along ecological axes, facilitating ecological speciation (Ellis et al. 2013). As such, limited dispersal could be a major determining factor in the high species richness of Succulent Karoo flora. To address these broad evolutionary questions requires much further investigation across multiple systems, with this study providing a small insight into the genetic structure of one Namaqualand Asteraceae species.

Chapter 4. Petal spot pigmentation and candidate regulatory genes

4.1 Introduction

The *Gorteria diffusa* petal spot is a heavily pigmented, three-dimensional elaboration of the petal epidermis that forms across fused petals of a single ray floret. It is unusually complex compared to the petal spots of several other daisy and eudicot species, which generally consist of pigment accumulation but not epidermal modifications (Thomas et al. 2009). The deep texture and rich colouration of petal spots in *G. diffusa* are created by an amalgamation of different specialised cell types that vary in colour due to pigmentation and cuticular elaborations. Interior cells *en masse* have an overall green appearance, but individually may be variably pigmented green, blue, purple or black. Aggregations of dark papillate cells are deep purple or black. Floral colouration is determined by the interactions of multiple components including cell shape that influences light reflection, pigment concentration, co-occurrence of several pigments, and the environment within which pigment accumulates, as well as the fundamental properties of the pigment itself (Brouillard 1983; Grotewold 2006). Anthocyanin pigmentation has an important role in the *G. diffusa* spot phenotype, although disentangling the precise contribution of anthocyanin to spot phenotype is difficult given the complexity of the spot.

Anthocyanins are a type of flavonoid, the glycosylated products of anthocyanidins. More than 30 naturally occurring anthocyanidins have been identified, six of which are considered common: cyanidin, delphinidin, malvidin, pelargonidin, peonidin, and petunidin (Corradini et al. 2011). These types of anthocyanin differ in the number and chemistry (H, OH, or OCH_3) of groups added to the basic chemical structure (Grotewold 2006). Generally, greater numbers of hydroxyl groups on the B-ring result in a more blue colouration, for example, *Iochrominae* species with blue flowers produce mainly delphinidin (three hydroxyl groups), red-flowered species produce pelargonidin (one hydroxy group), and the orange-red flowered species produces cyanidin (two hydroxy groups) (Larter et al. 2018; Tanaka et al. 2008). Anthocyanins are stored in the cell vacuole and, due to glycosylation, anthocyanins are less reactive and more water soluble than anthocyanidins (Corradini et al. 2011). Typically, hydroxyl groups (-OHs) of the glycosyls of anthocyanins are acylated by organic acids (aliphatic or aromatic acids), producing acylated anthocyanins in a process termed anthocyanin glycosyl acylation (Osawa 1982 in Zhao et al. 2017). Anthocyanins are often stored in plant vacuoles in the acylated form (Nakayama et al. 2003). The reactivity of anthocyanins are influenced by these factors, and acylation has been shown to enhance the stability of acylated anthocyanins - contributing to the stable colouration of flowers and fruits (Baublis et al. 1994; Teh and Francis, 1988; Zhao et al. 2017). The type, number, and acylation sites of the acyl groups varies; as such, glycosyl acylation creates much diversity within anthocyanin molecules (Andersen and Jordheim 2006 in Zhao et al. 2017). A single anthocyanin can contain different types of acyl groups simultaneously. The violet petals of *Lobelia erinus*, for example, are pigmented by Lobelinin A which contains one coumaryl, one malonyl and two caffeyl groups (Kondo et al. 1989). One plant organ can also contain many anthocyanins, with *Ajuga reptans* flowers producing diacylated and triacylated cyanins, and triacylated delphinins (Terahara et al. 2001).

The vacuolar pH can influence the perceived colour of anthocyanin pigmentation. Vacuole acidification in petunia flowers produces red petal colouration, and flower colour shifts to blue in mutants that prevent the hyperacidification of vacuoles (Faraco et al. 2014). Similarly, in *Ipomoea nil* blue colouration is enhanced by higher pH levels in the vacuole (Fukada-Tanaka et al. 2000). Acylation

also influences resistance to changing pH, with unacylated anthocyanins less resistant to increased pH values than acylated anthocyanins (Delgado-Vargas et al. 2000). At its most basal level, the factors influencing floral colouration derive from the composition of the anthocyanin molecule and its interaction with cellular features.

G. diffusa petal spots are pigmented with anthocyanins that are cyanidin derivatives, in the subset of morphotypes previously investigated (Walker 2012). *G. diffusa* ray florets are darkly pigmented on the abaxial side, and it is currently unclear whether the composition of anthocyanins is the same between petal spots and abaxial pigmentation. In *Clarkia gracilis* petals have a purple background colouration and a deep red/ purple petal spot. The anthocyanidins cyanidin and peonidin are only found in *C. gracilis* petal regions that contain spots, whereas malvidin is found throughout the petal in spotted and plain regions. Spot-specific anthocyanin pigments were derived from a different branch of the anthocyanin pathway than other floral pigmentation, implying that differential regulation at these branch points could be important for spot formation (Martins et al. 2013). An R2R3 MYB transcription factor (CgMyb1) was found to regulate spot development. *CgMyb1* has a spatially restricted expression domain and activates transcription of the gene encoding an anthocyanin synthesis enzyme (*Dfr2*) only in specific petal regions, confining peonidin and cyanidin pigment production to petal spots (Martins et al. 2017).

Within Asteraceae, R2R3 MYB genes involved in anthocyanin synthesis regulation have been identified in species including *Gerbera hybrida* (Elomaa et al. 2003; Laitinen et al. 2008) and *Chrysanthemum morifolium* (Liu et al. 2015a). Across a range of phylogenetically diverse species, the transcription of genes encoding anthocyanin synthesis enzymes is regulated by a trimeric transcription activation complex (MBW complex) comprised of an R2R3 MYB, bHLH, and WD40 transcription factor (Albert et al. 2011; Carey et al. 2004; Cone et al. 1986; Gonzalez et al. 2008; Goodrich et al. 1992; Lin-Wang et al. 2010; Ludwig et al. 1989; Paz-Ares et al. 1986, 1987; Quattrocchio et al. 1993, 1999; Ramsay and Glover 2005; Schwinn et al. 2006; Spelt et al. 2000). The MYB and bHLH proteins have sequence-specific DNA binding activity, while the WD40 protein provides a scaffold enabling bHLH and MYB protein-protein interactions. Within the complex, activity of the R2R3 MYB requires interaction with the bHLH partner. The bHLH is thought to stabilise the MYB protein and enhance activation of anthocyanin synthesis genes that contain a conserved *cis*-regulatory element (Grotewold 2006; Hernandez et al. 2004). The bHLH protein interacts with both the MYB and WD40 proteins, whereas the MYB component only interacts with the bHLH. As such, the bHLH protein is more constrained evolutionarily than the MYB because mutations in the former are more likely to disrupt functioning of the whole complex. Therefore, the MYB component of the MBW complex is thought to confer the greatest developmental specificity (Ramsay and Glover 2005). Interestingly, in *A. thaliana* the constituent proteins of the MBW complex controlling trichome differentiation have been shown to form dimers, with competitive binding of the MYB and WD40 to the bHLH; the two types of dimer can regulate expression of different genes (Pesch et al. 2015). A recent study in *Antirrhinum majus* suggests that this may also apply to transcription factors regulating floral anthocyanins, demonstrating that the complex of Rosea 1 (MYB) and Delila (bHLH) most effective at activating anthocyanin synthesis may lack the WD40 protein (Albert et al. 2020).

Analyses in several species indicate that the strict spatial control of anthocyanin pigment expression, necessary for spot formation, is often achieved through transcriptional regulation. This is an important mechanism for producing diverse floral pigmentation patterns in both monocots and eudicots. Within

Antirrhinum majus, regulation of anthocyanin-pigmented venation is controlled by a bHLH protein expressed in the epidermis and an R2R3 MYB transcription factor with circum-vein expression (VENOSA). The venation phenotype is restricted to regions where the bHLH and MYB spatially overlap – enabling complex formation and activation of anthocyanin production. Supporting this, the overexpression of *VENOSA* in unpigmented epidermal tissue (where the bHLH is expressed) leads to anthocyanin production. This suggests that pigment restriction is not due to repressor proteins that inhibit either *VENOSA* expression or VENOSA protein activity in unpigmented regions, but is the result of overlapping expression of bHLH and MYB proteins (Shang et al. 2011). In the petal lobes and corolla tube of *A. majus* the MYB genes *ROSEA1* and *ROSEA2* regulate anthocyanin synthesis. The intense colouration of the flower results from combined action of VENOSA, ROSEA1, and ROSEA2 activating anthocyanin biosynthesis in various petal regions (Schwinn et al. 2006). Similarly, vein-associated anthocyanin pigmentation in the corolla tubes of *Petunia* is regulated by the R2R3 MYB transcription factor DEEP PURPLE and different R2R3 MYB transcription factors regulate anthocyanin production across the corolla (An2) and in anthers and corolla tubes (An4). A fourth MYB, *PURPLE HAZE,* controls light-regulated accumulation of anthocyanin on the bud abaxial petal surface (Albert et al. 2011; Gerats et al. 1985; Gerats et al. 1984). In contrast to *Petunia* and *A. majus*, 'splatter' spot patterns on the tepals of *Lilium* spp. are independent of vein position and appear early in flower development. *LhMYB12-Lat* regulates the splatter pigmentation, while the other allele of the *LhMYB12* gene is not involved in spot regulation, instead activating background anthocyanin pigmentation in the tepals, filaments, and styles. *Lilium* species also have raised spots that form increased numbers of epidermal and parenchyma cells relative to the rest of the petal, in addition to pigmentation. The anthocyanin pigment within these spots is regulated by the transcription factor LhMYB6 (Yamagishi et al. 2010). Light-exposed surfaces of *L. regale* flower buds develop anthocyanin pigmentation regulated by another R2R3 MYB, LrMYB15 (Yamagishi 2016). Evidently, multiple *R2R3-MYB* genes often operate in a single species to regulate anthocyanin production, with the distribution of pigmentation determined by the spatial and temporal distribution of R2R3 MYB allele or gene transcription.

Previous work determined that cyanidin-3-glucoside is the anthocyanin pigmenting the petal spots of several *G. diffusa* morphotypes. A candidate petal spot regulator (*GdMYB8*) was identified through a candidate gene approach and comparative transcriptomics. Phylogenetic analysis placed this gene within the subgroup 6 R2R3 MYBs. R2R3 MYB subgroups are defined by conserved amino acid sequence motifs, and subgroup 6 MYBs function in anthocyanin regulation across many systems (Petroni and Tonelli 2011; Stracke et al. 2001). This chapter aims to determine whether the type of anthocyanin pigmentation is homogenous across different regions of *G. diffusa* ray floret petals, or if anthocyanin within petal spots has a different composition from abaxial pigmentation. Potential regulators of spot anthocyanin production are investigated and identified, expanding on previous work. The phylogenetic relationships between these regulators and other Asteraceae subgroup 6 MYB genes are assessed. Gene expression patterns are characterised to determine whether candidate genes are upregulated in petal spots, supporting a role in petal spot developmental regulation. These analyses are conducted across three *G. diffusa* morphotypes, two of which are thoroughly phenotypically characterised, with the natural variation present in floral phenotype contributing to the robustness of genetic findings. The suitability of these morphotypes for future evolutionary analyses was assessed and the developmental genetic data presented here will, hopefully, be expanded to a comparative evolutionary framework once the necessary tools and information are available to conduct evolutionary developmental work within the system.

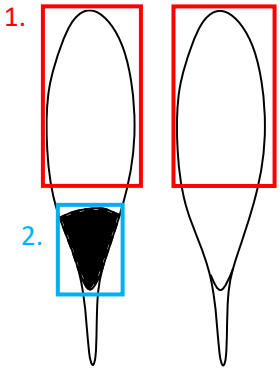
4.2. Methods

4.2.1 Phenotypic measurements of Spring and Cal morphotypes

Samples were collected during fieldwork in the Northern Cape of South Africa in 2018. Sample collection, preparation, and phenotypic measurements were conducted as described in Section 3.2.1. Phenotypic measurements of individuals (n = 12 per population) were taken from 4 populations within the Cal morphotype range. These population are considered 'pure' Cal (photograph in Fig 4.7) and are situated in the central morphotype range and not toward the peripheries close to contact zones with other morphotypes. Equivalent Spring individuals were sampled (n = 12 per population) from 6 'pure' Spring populations. Measurement of phenotypic traits was automated in R (Section 3.2.5) by Boris Delahaie and the variables measured are listed in the table below. A principal component analysis was conducted on the trait measurements listed below (Table 4.1) in the R package nsprcomp (Sigg and Buhmann, 2008) and visualised in the factoextra package (Kassambara and Mundt, 2017). The relative PC loadings of each trait, and mean trait values for Cal and Spring are listed in Supp. Table 3.

4.2.2 Anthocyanin extraction

The anthocyanin extraction procedure is detailed in Section 2.8. Several different tissue types were sampled from the morphotypes Spring, Cal, and Stein. Individual ray florets were plucked from mature capitula and dissected with a razor blade. The same segments from different individual ray florets and capitula within a plant were pooled and then snap frozen in liquid nitrogen. Two types of ray floret segments were dissected, depending on the morphotype phenotype and ray floret type (Fig 4.1).



Morphotype and Tissue	Sample Size	Segment Type
Cal Plain	6	1
Cal Spot	7	2
Spring Plain	6	1
Spring Spot	6	2
Spring Mark	5	2
Stein Plain	7	1

Figure 4.1 Schematic of the tissues dissected for HPLC analyses. The adaxial side of generic spotted and plain ray florets are shown. The segments of ray florets dissected are indicated in the diagram: 1. 'Plain' and 2. 'Spot/ Mark'. Spot refers to a complex petal spot composed of multiple cell types, while mark is a simple spot composed only of black pigment.

Previous analysis indicated that chlorophyll is present in the ray florets of Spring (dissected spot: 311/g ± 32, plain ray floret 141/g ± 2.8) and Cal (dissected spot: 1019/g ± 54, plain section of ray floret 914/g ± 55) (Walker 2012). As such, the absorbance of each pigment extraction (Fig 4.2) was measured at A530 (the peak of absorption of anthocyanin) and A657 (the peak of absorption of degradation products of chlorophyll in acidic methanol). After correcting for the dilution factor, the equation $A = A_{530} - 0.25A_{657}$ was used to calculate the total relative anthocyanin content (Mancinelli 2020). This value (A) was then divided by the fresh weight of the sample to give (A_2). Approximate absolute anthocyanin concentrations were calculated using the equation: concentration=(A_2/34) x 484.83 (Airoldi et al. 2019), where 34 is the millimolar extinction coefficient (Gerats et al. 1982) and 484.83 is

the molecular weight of cyanidin 3-glucoside. Calculations were completed in excel and graphical representation of the results were completed in R (packages:ggplot2 (Wickham 2016)). Statistical modelling was completed using R packages multcomp (Hothorn et al. 2008), ggfortify (Horikoshi and Tang 2018; Tang et al. 2016), and nlme (Pinheiro et al. 2012). A linear mixed model and Tukey honest significant differences test were used to determine whether differences in anthocyanin content between tissues were statistically significant. Individual plant sampled was added as a random effect to account for variation resulting from factors specific to individuals. The anthocyanin content data were transformed so that the data did not violate any of the assumptions of the linear mixed model.

Figure 4.2 A subset of the anthocyanin extractions in acidic methanol. Dilutions of these samples were used to quantify overall anthocyanin concentrations. Labels indicate the sample and weight (from left to right): Cal plain (68mg), Cal spot (52mg), Cal plain (107mg), Cal spot (40mg), Spring spot (28mg), Spring plain (146mg), Stein plain (80mg), Stein plain (110mg).

4.2.3 Anthocyanin HPLC-MS analysis

Tissues dissected for HPLC were from the same morphotypes and ray floret segments as that used for overall anthocyanin quantification. The top of the Spring spotted ray floret petals was added as an additional sample type. From visual assessment these petal sections do not appear to be pigmented by anthocyanin. The sample size for each type of tissue was n = 3. Tissue preparation is detailed in Section 2.8, with the modification that samples had 1ml of acidic methanol (1% (v/v) 1M HCL) added and were shaken overnight only once, rather than this process being repeated. Pigment extractions were put on ice and sent to John Innes Centre, Norwich and stored at -20^{o}C until the analysis. HPLC-MS was conducted by Lionel Hill. The samples were diluted two-fold with ddH_2O, centrifuged, and transferred to 200µl glass inserts for the analysis. The samples were analysed on a Prominence/Nexera UHPLC system attached to an ion-trap ToF mass spectrometer (Shimadzu). Separation was on a 100 × 2.1mm 2.6µ Kinetex EVO column (Phenomenex) using the following gradient of acetonitrile versus 1% (v/v) formic acid in water, run at 0.5ml/min and 40°C:

time (minutes)	% acetonitrile
0.01	2
0.50	2
5.00	10
17.00	30
25.00	90
25.80	90
26.00	2
30.10	2

Detection was by UV/visible absorbance and positive mode electrospray MS. The diode array detector collected spectra from 200 - 650nm at 6.25 spectra/sec, with a time constant of 0.08 secs. The MS collected spectra from *m/z* 220 - 2000, with automatic sensitivity control set to a target of 70% optimum base peak intensity. It also collected automatic (data dependent) MS2 spectra from *m/z* 50 - 2000 with a fixed ion accumulation time of 20msec, an isolation width of *m/z* 3.0, 50% collision energy and 50% collision gas. After two spectra had been collected for a precursor ion, it was ignored for 3 secs in favour of the next most abundant ion. Spray chamber conditions were 250°C curved desorbation line, 300°C heat block, drying gas 'on', and 1.3 l/min nebuliser gas. The instrument was calibrated immediately before analysis, using sodium trifluoroacetate cluster ions, according to the manufacturer's instructions. Data output was summarised in excel and graphs were made in R (package: ggplot2 (Wickham 2016)).

4.2.4 Characterising *G. diffusa* subgroup 6 R2R3 MYB genes

A promising candidate for petal spot pigmentation, *GdMYB8* (here termed *GdMYB8a*), was previously identified through a 454 transcriptome analysis conducted in the Spring morphotype. This gene was shown to have spot specific expression patterns (Mellers 2016; Walker 2012). Two additional *GdMYB8* homologues, *GdMYB8b* and *GdMYB8c*, were identified initially through genome walking (Section 2.3.3). A more recent RNA-seq analysis (Kellenberger unpublished) led to the identification of a fourth homologue, *GdMYB8d*. The full-length sequences of these genes (including 3'UTRs) were obtained through PCR and 3'RACE (Section 2.3.5). The primers used (Appendix 2) were designed based on transcriptome sequences and predicted conserved regions with *GdMYB8a*. To characterise the variation in these genes, the cDNA and gDNA sequences for each of the 4 candidate genes were obtained through PCR for multiple individuals per morphotype (with the exception of Stein *GdMYB8b*). As Spring was the focal morphotype for this analysis, more stringent characterisation was conducted. In Spring, for each individual, 5 identical PCR reactions were conducted per gene to ensure genetic variation was captured. All amplicons were subsequently cloned (Section 2.4) and sent for Sanger sequencing. Sequencing data was formatted and analysed in Geneious Prime and Benchling (Biology Software).

4.2.5 Building an Asteraceae subgroup 6 R2R3 MYB amino acid phylogeny

Asteraceae MYB subgroup 6 protein sequences were obtained through BLAST analysis in Genbank (Benson et al. 2012), the sunflower (*Helianthus annuus* L.) genome (https://www.sunflowergenome.org/, Badouin et al. 2017) the lettuce (*Lactuca sativa*) genome (Lettuce Genome Resource, https://lgr.genomecenter.ucdavis.edu/, Reyes-Chin-Wo et al. 2017), and through a literature search for papers characterising subgroup 6 MYBs in Asteraceae (Yue et al., 2018). Hypothetical proteins and duplicates of the same gene product were removed. The *G. diffusa* GdMYB2 (Thomas 2009), a subgroup 9 R2R3 MYB, was added as an outgroup. *G. personata* gDNA was extracted from samples taken in the field and *GdMYB8a-c* were amplified through PCR. Introns were predicted by aligning the *G. personata* gDNA sequence with *G. diffusa* coding sequences, removing predicted intron segments, and translating these DNA sequences into amino acids – no premature stop codons were found. The phylogenetic analysis derives from RAxML analysis of amino acid sequences. The amino acid sequences were aligned with mafft (v 7.429) (Katoh et al. 2002) using default parameters. The model of molecular evolution was selected using PartitionFinder (v2.1.1) (Lanfear et al. 2017) testing all amino acid models available assuming one data block. The phylogenetic tree was inferred using maximum-likelihood optimality criterion with RAxML-NG (v.0.9.0) (Kozlov et al. 2019) while

calculating bootstrap branch support. The phylogenetic tree was visualized by FigTree (v1.4.4) and edited in inkscape (v1.0.2) (Inkscape 2020). Final alignments and tree building were conducted by Qi Wang.

4.2.6 Examining expression levels of *GdMYB8* genes

The expression levels of *GdMYB8* genes within *G. diffusa* ray florets were tested using qRT-PCR in Spring, Cal, and Stein. Ray floret tissue was collected at two developmental stages during spot development, based on the developmental characterisation of the Nieuw morphotype by Thomas et al. (2009). Spring and Cal were examined under the microscope at different developmental time points to determine whether the timing of spot development was roughly equivalent in each morphotype. At developmental stage 1 the spot is a small dark patch on an otherwise yellow/ green ray floret, by developmental stage 2 specialised cell types are forming and abaxial pigment is initiating. Stage 1 was defined as the point when ray florets had between the same height and x1.5 the height of developing disc florets and stage 2 when the ray florets were at least double the height of developing disc florets, but not yet fully mature (Fig 4.3). In Cal each ray floret was dissected into two latitudinally: a non-spotted (plain) top segment and a spotted bottom segment. In Spring, the proportion of the ray floret that the spot occupies is highly variable and at the first developmental stage the early spot is enclosed by peripheral petals of the ray floret. Consequently, it is very difficult to accurately dissect these ray florets into spotted and non-spotted segments. The comparison in Spring was instead done between whole plain ray florets and whole spotted ray florets. Diagrams indicating the tissue segments used are in Fig 4.16 and 4.17. Spring individuals with a 'mark' (simple spot at the base of the plain ray floret) above a certain size were excluded, as this could confound the 'spotted' and 'non-spotted' comparison. Only non-spotted Stein plants were used for expression analysis and whole ray florets were sampled. As some individual Stein plants have been observed to switch on and off spot development, every ray floret was carefully checked for spots during tissue collection. Each biological replicate consisted of 3 plants and samples were pooled within biological replicates, all tissue collection was conducted at 9:15 - 10:30am to account for the potential confounding effects of circadian rhythm on gene expression. Samples were prepared and RNA extracted using the procedures outlined in Sections 2.2.3 and 2.2.4.

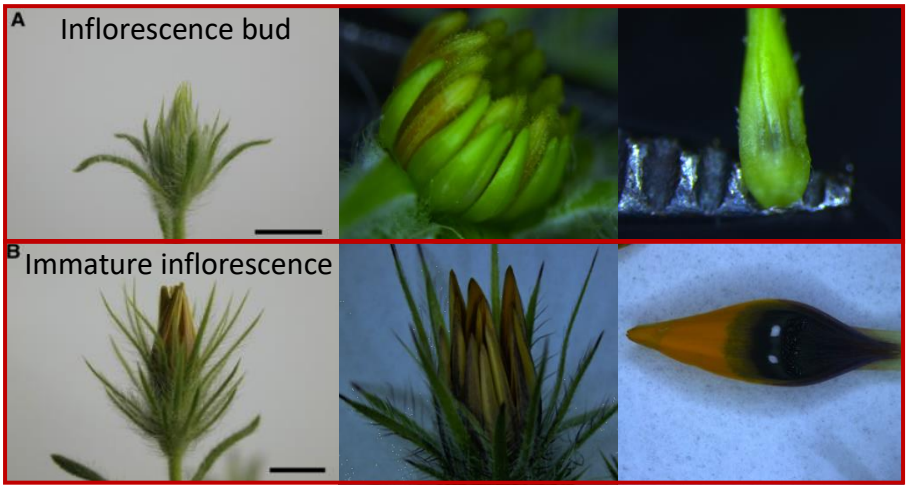


Figure 4.3. The developmental stages used for qRT-PCR. The Spring morphotype is pictured at developmental stage 1 (A) and stage 2 (B). Scale bars = 1cm.

The full qRT-PCR procedure is detailed in Section 2.7. Designing qRT-PCR primers specific to each *GdMYB8* gene was challenging due to the similarity between them. An alignment of Asteraceae MYBs from different subgroups was used to exclude regions conserved across MYB genes. The regions deemed potentially suitable were then included in an alignment with all *G. diffusa MYB8* sequences so that sites consistently different between genes could be identified. Suitable segments for primer design were found at the 3' end of the gene and into the 3' UTR (isolated through 3' RACE, detailed in Section 2.3.5). At least one qRT-PCR primer per pair was designed to be specific to one *GdMYB8* gene. Specificity of amplification by primer pairs was assessed for *GdMYB8a/b/c* through PCR reactions. A serial dilution of vectors containing each *GdMYB8* gene as a template was used to determine if amplification could be achieved with primers designed for a different homologue (example in Fig 4.4). *GdMYB8d* was considered divergent enough that this specificity check was not required. Following qRT-PCR the PCR products were sent for Sanger sequencing and the sequences analysed to check that only the intended gene product was amplified.

GdEF-2 (Elongation Factor 2) was used as a reference gene (Section 2.7.3), but primers were originally designed from the Spring morphotype only. *GdEF-2* was here isolated (primer in Appendix 2) in Cal and Stein to determine whether or not the primer sequences were conserved between morphotypes. Primers to amplify *GdEF-2* were designed from the Spring gene sequence previously obtained by Mellers (2016). *GdEF-2* fragments were amplified in multiple Cal and Stein individuals through PCR, cloned, and sequenced (Section 2.3 and 2.4). There were consistent polymorphisms in the forward primer (i.e. Stein and Cal had identical sequences, but this differed from Spring), so the *GdEF-2* forward qRT-PCR primer was redesigned (Appendix 2) and primer efficiency retested. Similarly, some of the *GdMYB8* qRT-PCR primers contained polymorphic sequences between morphotypes, these were redesigned to make them morphotype specific, primer efficiency was tested and specificity checks conducted as outlined above. Programs used for statistical tests and graphical representations of the data are detailed in Section 2.7.4.

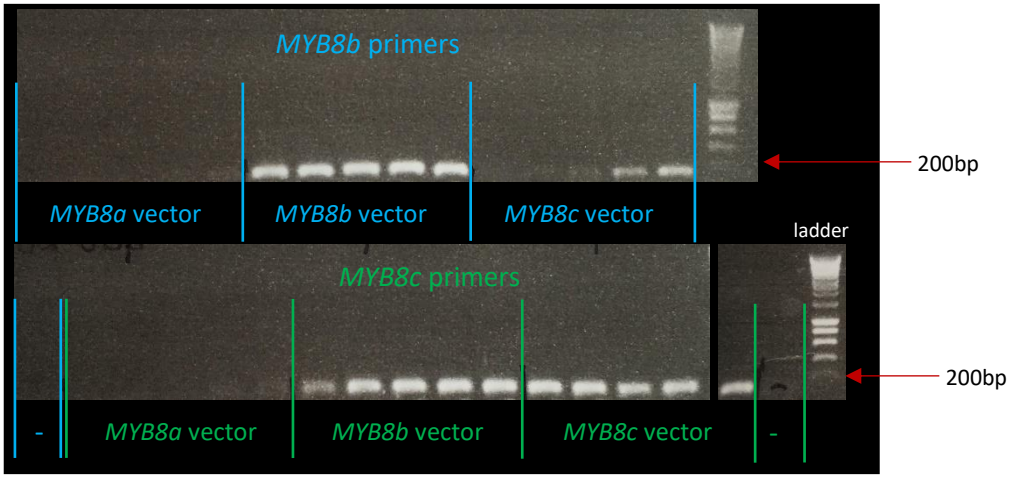


Figure 4.4. Example of the qRT-PCR primer specificity test. Each of the five wells per vector represent a different concentration of template plasmid (from left to right: 0.001ng/µl, 0.01 ng/µl, 0.1 ng/µl, 1 ng/µl, 100 ng/µl). Lettering colour indicates PCR reactions using the *MYB8b* primer pair (blue) and *MYB8c* primer pair (green). In this case the *MYB8b* primers were considered sufficiently specific (not amplifying the *MYB8a* or *MYB8c* vector at concentrations > 0.1 ng/µl), whereas the *MYB8c* primers were not – these were discarded, and new primers designed. The expected length of amplicons were 170bp for *MYB8b* primers and 152bp for *MYB8c* primers.

4.3 Results

4.3.1 Selecting *G. diffusa* morphotypes for comparative analyses

Field observations and phenotypic differences

The natural phenotypic variation between floral morphotypes provides a useful comparative framework for understanding spot development, here used to improve robustness of genetic characterisation. Spring was the focal morphotype and the two additional morphotypes were Stein, which contains non-spotted individuals, and Cal that has complex spots arranged in a bullseye pattern. Cal petal spots are raised but lack the white highlight cells and papillae found within Spring. All of these morphotypes have been karyotyped and are diploid with 10 chromosomes in total (Thomas 2009).

Morphotypes were also assessed for suitability for comparative evolutionary analyses, although this is beyond the scope of the current study. Pollinator behavioural assays demonstrated that the Cal floral phenotype induces different behavioural responses in *M. capensis* pollinators compared to Spring capitula (Ellis and Johnson 2010). There are Cal populations geographically adjacent to Spring populations but no contact zones are known, and phenotype within Cal is relatively consistent based on quantitative spot trait measurements in one Cal population (Ellis et al. 2014). The morphotype Stein encompasses spotted and non-spotted individuals. From observations in the field, it appears that certain populations are predominantly comprised of non-spotted individuals – while other populations are mainly spotted, but further characterisation is required. Stein seeds collected from several populations were grown in the glasshouse and some individuals were found to have both spotted and non-spotted capitula (Figure 4.6). Consequently, Stein 'non-spotted' individuals cannot be used as non-spotted *G. diffusa* representatives in evolutionary analyses because it is, currently, unclear whether or not they are capable of switching on spot production.

Floral traits of Cal and Spring form discrete phenotypic clusters

Individuals were sampled from multiple populations of Cal and Spring morphotypes. Phenotypic measurements of floral traits were taken, and a principal component analysis of ray floret traits was conducted. The principal components included trait measurements relating to spot size and ray floret colouration (detailed in Table 4.1). The first two principal components (PCs) cumulatively explained 77.7% of the variation (PC1 67.5%, PC2 10.2%) in floral traits. The trait variable loadings for each PC and mean trait measurements for Spring and Cal are listed in Appendix 4 (Supp. Table 3). As illustrated in Figure 4.5, there is phenotypic variation within each morphotype, but the two morphotypes are clearly differentiated by floral traits along PC1 – clustering into two discrete groups according to morphotype.

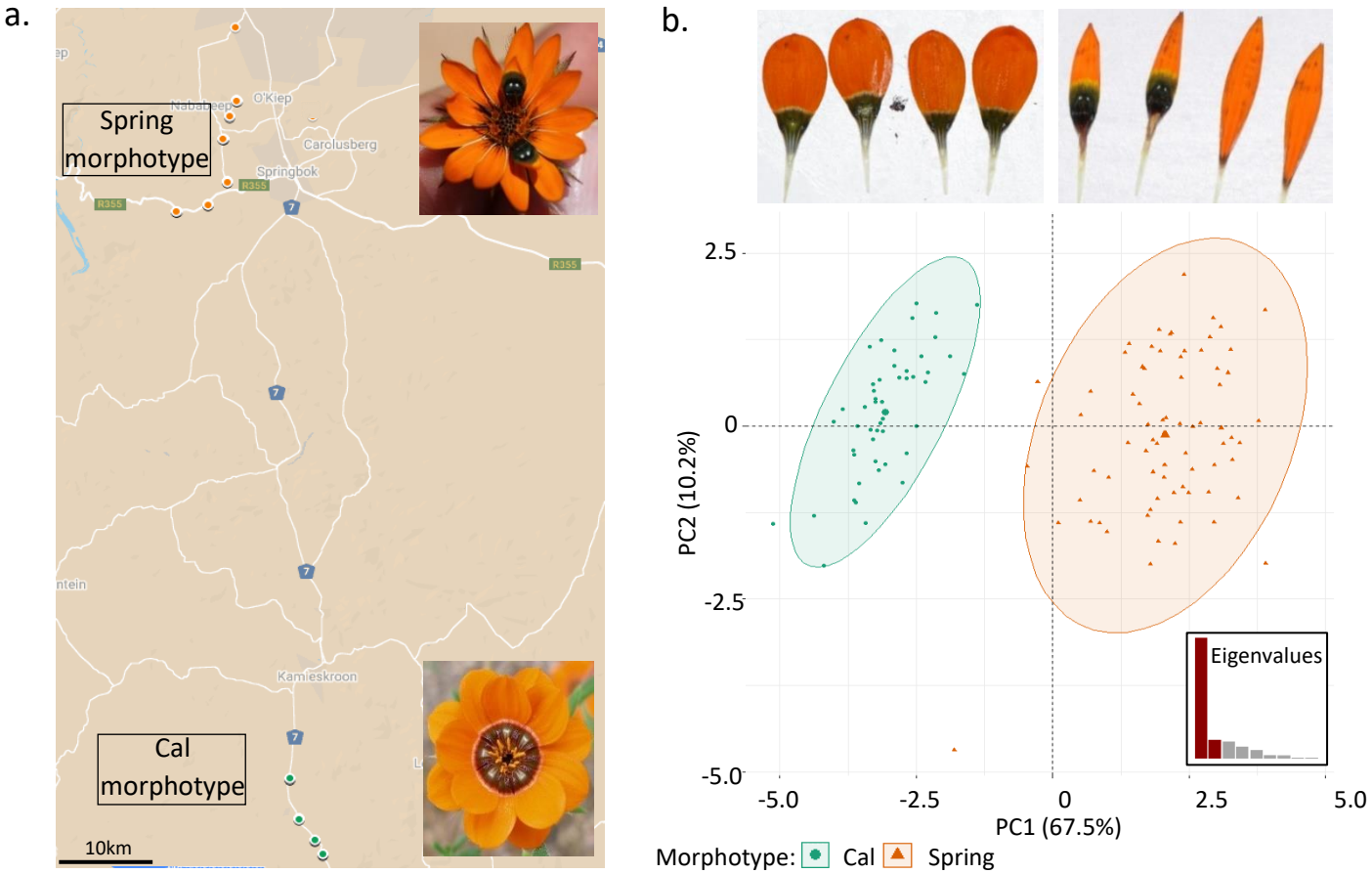


Figure 4.5. Comparison of floral phenotypes between the morphotypes Spring and Cal. a) The locations of the populations that individuals were sampled from for phenotypic analysis (n_{total} = 120, n_{Cal} = 48, n_{Spring} = 72, $n_{perpopulation}$ = 12). Orange dots indicate Spring populations and green dots indicate Cal populations – an image of a typical *G. diffusa* capitulum corresponding to each morphotype is pictured on the map. b) Examples of the photographs used for automated phenotypic trait measurements are presented. The first two principal components from the PCA of floral phenotypic measurements are represented graphically (PC1 = 67.5%, PC2 = 10.2%), with 95% confidence ellipses. The eigenvalue graph demonstrates the percentage of variation accounted for in the first 10 eigenvectors and red bars indicate the eigenvalues represented in the PC scatterplot.

Trait measurements
Proportion of ray florets within a capitulum that are spotted
Red values (from red-green-blue extracted values) of the plain segment of the spotted ray floret
Green values (from red-green-blue extracted values) of the plain segment of the spotted ray floret
The length of the plain segment of the spotted ray floret (from the top of the spot to the tip of the ray floret)
Blue values (from red-green-blue extracted values) of the plain segment of the spotted ray floret
Mean brightness value on the plain segment of the spotted ray floret
Aspect ratio of the spotted ray floret
Aspect ratio of the spot
The length of the plain segment of the spotted ray floret as a proportion of the total ray floret length
The area of the spot as a proportion of the total spotted ray floret area

Table 4.1 Descriptions of the variables used in a PCA analysis comparing the phenotypic traits of Cal and Spring spotted ray florets.

4.3.2 Ray florets are pigmented by cyanidin 3-glucoside

Most ray floret tissues contain anthocyanin

All three focal morphotypes (Spring, Stein, Cal) have purple or black pigmentation on the abaxial side of the ray floret. The extent of this pigmentation varies between individual plants within glasshouse grown individuals. Cal and Stein tend to have consistently darker pigmentation extending over more of the petal area than Spring, while Spring abaxial pigmentation was more variable (personal observation). The abaxial colouration was found to vary in individuals growing in wild populations of *G. diffusa* (Fig 4.6c, d), but Cal and Stein abaxial pigmentation was not characterised in the field. Fig 4.6d gives examples of wild Cal individuals and Fig 4.5.bii shows glasshouse grown Stein. The abaxial side of the raised spotted ray floret petals in Spring do not appear to be pigmented by anthocyanin based on colouration.

Petals of ray florets from mature capitula of the focal morphotypes were dissected into spotted and non-spotted segments (illustrated in Fig 4.7a), pooled according to morphotype and tissue type, and an anthocyanin extraction in acidic methanol was performed. Anthocyanin content was quantified and compared between tissue types (Fig. 4.7b). All tissue types contained anthocyanin. Spring spots had the greatest range of anthocyanin content and a significantly higher anthocyanin content than any other tissue type (comparisons to Spring spotted tissue: Cal plain: $z = 3.38$, $p = 0.009$; Stein plain: $z = 4.58$, $p < 0.001$; Cal spot: $z = 4.66$, $p < 0.001$; Spring mark: $z = -7.02$, $p < 0.001$; Spring plain: $z = 9.00$, $p < 0.001$). This is perhaps due to high anthocyanin content within the swollen epidermal cells of the Spring papillae (Thomas et al. 2009). Cal spots had a similar anthocyanin content to Cal and Stein plain ray floret petals. Spring plain ray floret petals contained significantly less anthocyanin than Cal plain ray floret petals ($z = -3.57$, $p = 0.005$). The anthocyanin content of each tissue was as follows (mean ± s.d.): Spring spots 1.05µg/mg ± 0.45, Cal plain 0.48µg/mg ± 0.34, Stein plain 0.36µg/mg ± 0.20, Cal spot 0.38µg/mg ± 0.09, Spring mark 0.28µg/mg ± 0.04, and Spring plain 0.19µg/mg ± 0.07. The Cal

spot anthocyanin content is consistent with that found by Thomas et al. (2009), who demonstrated that Cal spot anthocyanin content was twice that of *Antirrhinum majus* petals.

An anthocyanin with a malonyl residue pigments petal spots

To determine the anthocyanin content of *G. diffusa* ray floret tissue, pigments were extracted from mature ray florets using acidified methanol (1% (v/v) HCL) and analysed by HPLC-MS. Compounds were detected by UV/ visible absorbance (200 - 650nm) and, subsequently, electrospray mass spectrometry (collecting spectra from m/z 220 - 2000) was used for peak assignment and further characterisation of the substances detected – through fragmentation of abundant ions. Samples were analysed from spotted and plain regions of ray floret petals from the morphotypes Cal, Spring, and Stein (Fig 4.7a).

Across all samples, the chromatogram showing UV absorbance detected a major peak at 5.059 mins. This was the only major peak in all of the samples that did not contain petal spot tissue. The compound mass was 449, with collision-induced dissociation (CID) of the $[M]^+$ ion at m/z 449 producing a base peak at m/z 287. This corresponds to aglycone cyanidin and the mass loss of 162 is consistent with loss of a glucose moiety (compound 1 in Table 4.3, peak 1 in Fig 4.8). Schütz et al. 2006 used an acidic methanol extraction for HPLC-MS on the Asteraceae species *Cynara scolymus* L. with a cyanidin 3-glucoside reference compound. The peak that they identified as cyanidin 3-glucoside had identical MS-MS analysis results as the major peak identified here, so we tentatively identify this substance as cyanidin 3-glucoside.

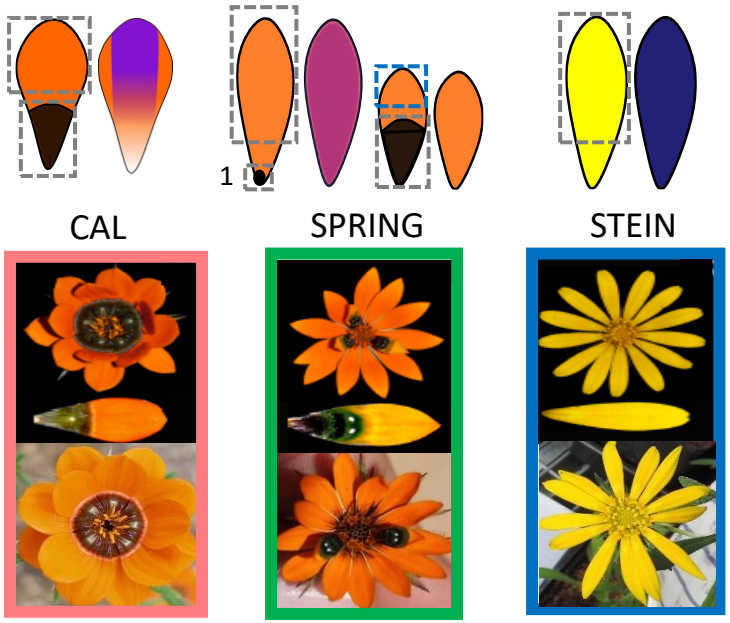
In all petal spot samples (Cal spot, Spring spot, Spring mark) there were two additional major peaks in the UV absorbance chromatogram. The first (compound 4, Table 4.3) had a retention time of 7.133 mins and the second (compound 5, Table 4.3) was detected at 8.283 mins. Both produced peaks at m/z 287 and m/z 449, indicative of a cyanidin glucoside. Compound 4 had a mass of 535 and a peak at m/z 491 – the mass loss of 44 is likely the loss of carbon dioxide from the terminal carboxylic acid group of malonate. Compound 4 had a mass of 549 and a peak at m/z 517, the mass loss of 100 is appropriate for a methylmalonate. There were several additional minor peaks in the UV absorbance chromatogram present in some of the samples, the physical properties of which are outlined in Table 4.3. All peaks were identified, through MS-MS, as cyanidin (though inconclusively in one case) and all but one contained a glucose moiety; the exception being a cyanidin with a pentose sugar moiety. In three of the compounds caffeate was detected, for example, one peak contained the characteristic MS peak of a cyanidin (m/z 287) with a compound mass of 611; the mass loss of 324 is appropriate for a glucose (m/z 162) and the remainder m/z 162 for a caffeate, with the long UV retention time in the HPLC analysis suggesting a hydrophobic decoration rather than two glucose or two caffeate moieties.

A rough estimation of the total relative anthocyanin content was calculated as the sum of all relative peak areas (corrected by weight) in a sample and was found to be highly consistent with the results of the anthocyanin quantification done with a larger sample size on the spectrophotometer. In the HPLC-MS analysis the top of the spotted ray floret 'Spring Top' in Spring was added as an additional sample, excluded from the previous anthocyanin quantification because it appeared unpigmented and so would likely be below the detection threshold of the spectrophotometer. HPLC-MS demonstrated that this tissue contained 92.3% - 98.4% less anthocyanin than the other tissues sampled. Cyanidins containing a malonyl group were more prevalent in spotted petal tissue compared to plain petal tissue across morphotypes and accounted for approximately 60% of the anthocyanins present within these samples (Fig 4.7c). The simple (just pigment with no cellular elaborations) spot on plain Spring ray

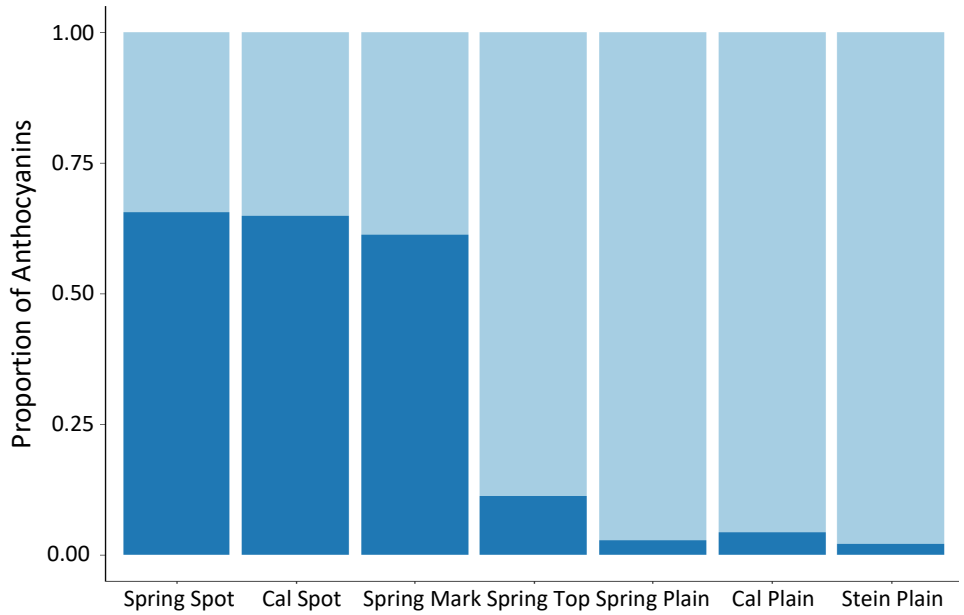
florets (‘Spring mark’) had a similar anthocyanin composition to the complex spots of Spring and Cal. Spring spots and Spring marks contained cyanidin glucosides with caffeate residues, accounting for approx. 3.6% and 1.6% of the anthocyanins in the samples, respectively. Anthocyanins containing caffeate also constituted 0.6% of the anthocyanins in plain Stein ray florets but were absent or present in trace amounts in all Cal samples, both spotted and plain. Overall, the complex spots of Spring contained the greatest diversity of anthocyanins, however, the overall quantity of anthocyanin analysed was greater within these samples. As such, it is possible that some of the compounds are present in other tissues and morphotypes but not in the detectable range of the HPLC-MS.

Peonidin glucoside coelutes with the isomer of cyanidin glucoside with a malonyl group (CyanGlcMal isomer) that was detected, so peaks cannot be easily distinguished. From the MS-MS data it also appears that peonidin glucoside may be present, but it is ambiguous. Based on MS peak areas, a semi-quantitative overview of which compound dominates (CyanGlcMal isomer or CyanPentose) indicated that peonidin glucoside might be relatively more prevalent in plain tissues across all morphotypes. Unfortunately, further consolidation is beyond the scope of the current analysis.

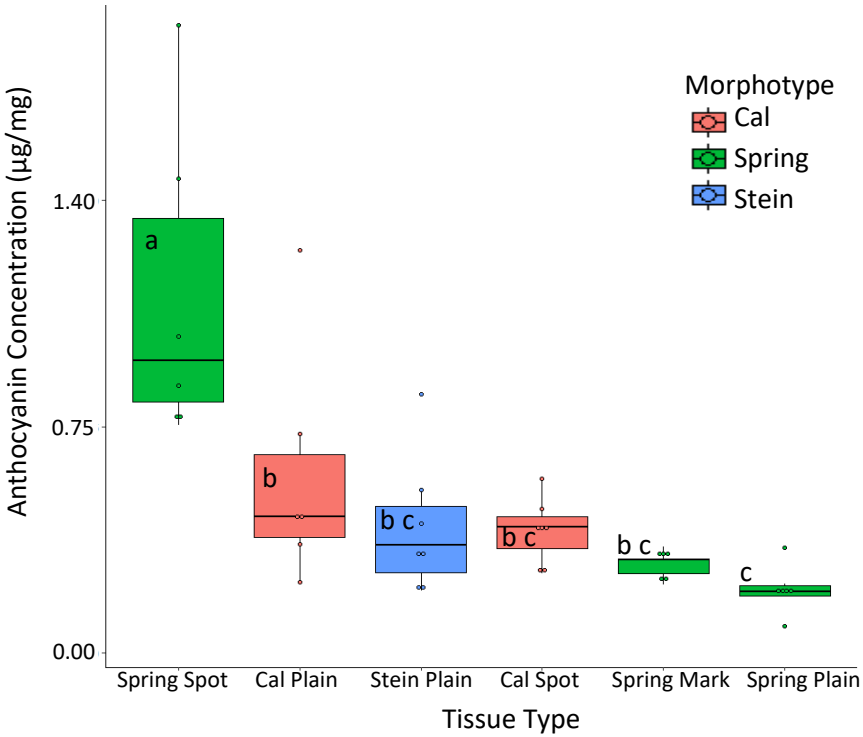
a.


1
CAL
SPRING
STEIN

ci.


Proportion of Anthocyanins
1.00
0.75
0.50
0.25
0.00
Spring Spot
Cal Spot
Spring Mark
Spring Top
Spring Plain
Cal Plain
Stein Plain

b.


Anthocyanin Concentration (µg/mg)
1.40
0.75
0.00
Morphotype
Cal
Spring
Stein
a
b
b c
b c
b c
c
Spring Spot
Cal Plain
Stein Plain
Cal Spot
Spring Mark
Spring Plain
Tissue Type

cii.

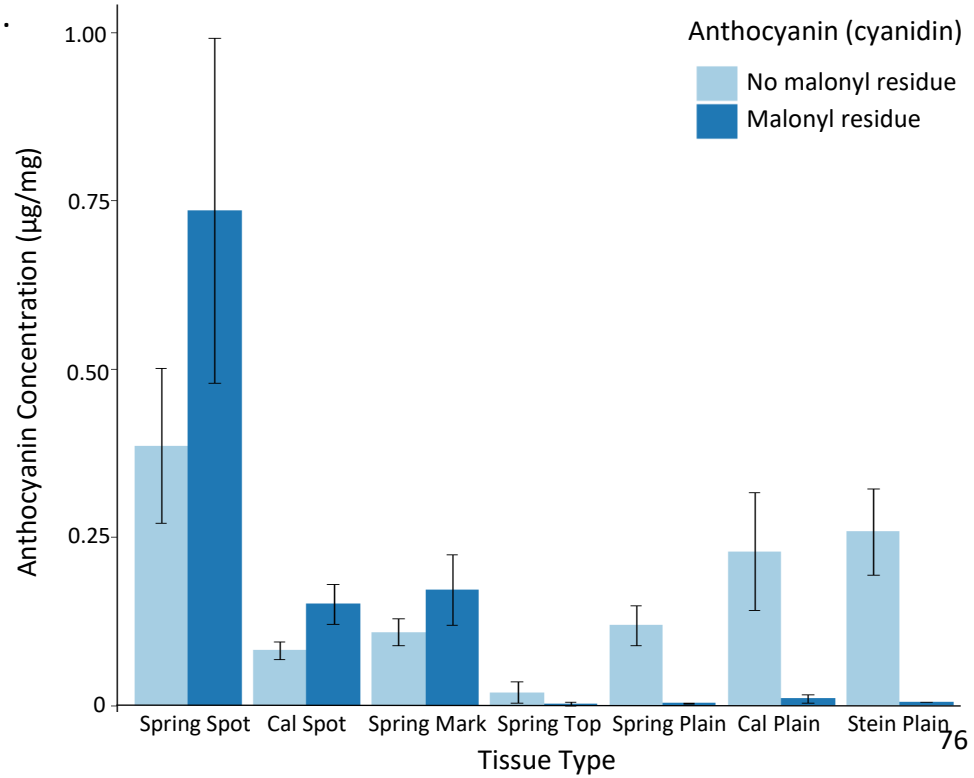

Anthocyanin (cyanidin)
No malonyl residue
Malonyl residue
Anthocyanin Concentration (µg/mg)
1.00
0.75
0.50
0.25
0
Spring Spot
Cal Spot
Spring Mark
Spring Top
Spring Plain
Cal Plain
Stein Plain
Tissue Type

Figure 4.7. (on previous page) Anthocyanin content of Cal, Spring, and (non-spotted) Stein morphotypes of *G. diffusa*. a) Schematics of typical inflorescences from each morphotype. The diagrams illustrate a typical ray floret adaxial (left) and abaxial (right) surface, with locations of anthocyanin indicated by black/purple colouration. Grey boxes indicate the tissues dissected for analyses illustrated graphically in b) and c). The blue box around the top of the Spring spotted ray floret indicates its use as an additional tissue type in c) analyses. '1' indicates the segment that represents 'Spring Mark', it is a small patch of pigment located at the base of plain Spring ray florets in some individuals. b) Overall approximate anthocyanin content for each tissue type (depicted in a), with the black line in each box indicating the median value and the whiskers 25/75% quantile +/- 1.5 *IQR, respectively. Individual data points are represented by black dots. Tissues that have anthocyanin concentrations which are significantly different from one another ($p \leq 0.05$) do not share letters. Sample size n = 5 - 7, where n represents pooled tissue from a single individual. c) Summary HPLC-MS analysis results. Approximate anthocyanin content for each tissue type is shown, grouped according to whether a malonate residue is present or absent. Sample size n = 3, where n represents pooled tissue from a single individual. ci) the proportion of anthocyanin that contains a malonate residue cii) the approximate anthocyanin concentration, error bars are +/- S.D.

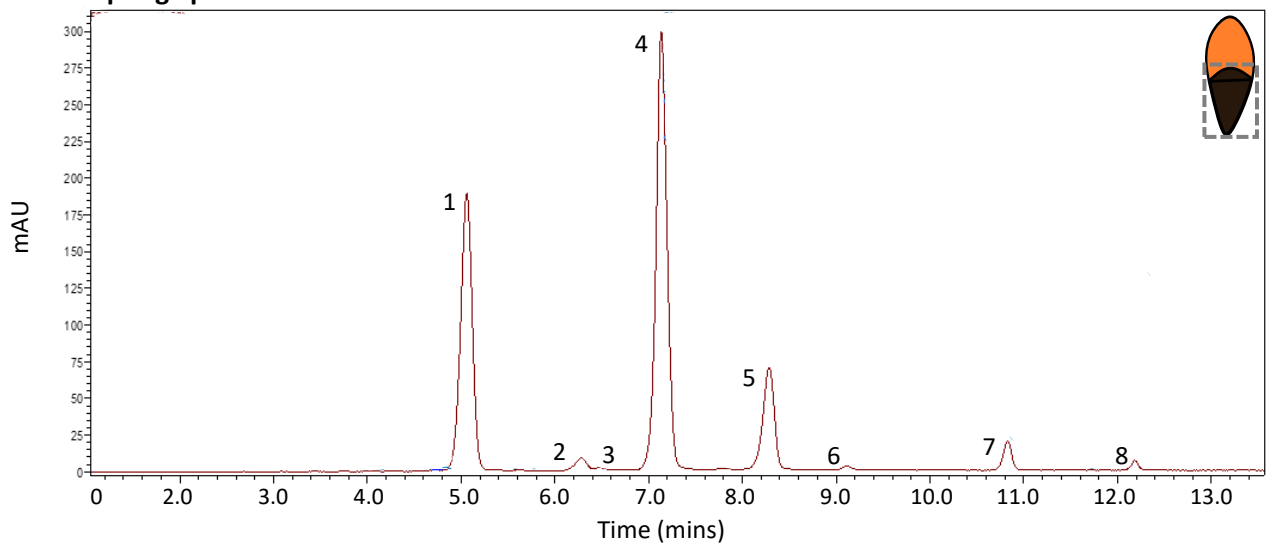
Compound	Spring Spot	Cal Spot	Spring Mark	Spring Top	Spring Plain	Cal Plain	Stein Plain
CyanGlc	0.3610 ± 0.0132	0.1331 ± 0.0139	0.1087 ± 0.0084	0.0244 ± 0.0011	0.1796 ± 0.0009	0.4551 ± 0.0083	0.3452 ± 0.0007
CyanGlcMal isomer	0.0156 ± 0.0008	0.0048 ± 0.0005	0.0038 ± 0.0004	0	0.0010 ± 0.0005	0.0010 ± 0.0004	0.0071 ± 0.0007
CyanPentose	0.0033 ± 0.0003	trace	0.0005 ± 0.0002	trace	trace	0.0014 ± 0.0007	0.0007 ± 0.0001
CyanGlcMal	0.5051 ± 0.0121	0.1676 ± 0.0121	0.1164 ± 0.0063	0.0014 ± 0.0006	0.0040 ± 0.0007	0.0196 ± 0.0068	trace
CyanGlcMeMal	0.1346 ± 0.0186	0.0693 ± 0.0021	0.0422 ± 0.0029	0.0005 ± 0.0004	trace	0.0035 ± 0.0019	0
CyanGlcCaf	0.0040 ± 0.0008	trace	0	0	0	trace	0.0021 ± 0.0004
CyanGlcCafMal	0.0266 ± 0.0039	trace	0.0028 ± 0.0006	trace	0	trace	trace
CyanGlcCafMeMal	0.0043 ± 0.0018	0	0.0013 ± 0.0002	0	0	0	trace
Total anthocyanins	1.0544 ± 0.4550	0.3761 ± 0.0945	0.2755 ± 0.0409	0.03*	0.1852 ± 0.0727	0.4810 ± 0.3445	0.3559 ± 0.1981

Table 4.2. Types of anthocyanin detected in *G. diffusa* ray floret tissue through HPLC-MS. Cyan = cyanidin, Glc = glucoside, Caf = cafeate residue, Mal = malonyl residue, MeMal = methylmalonyl residue, Pentose = pentose sugar. Each value represents the approximate anthocyanin concentration (µg/mg) of each compound (± s.e., n = 3), calculated by multiplying the proportion of anthocyanin the compound represents with the total anthocyanin content (Section 4.3.2). Trace is used where only 1/3 samples contained the compound and the mean relative quantity was <0.0006. *Estimated from total peak areas relative to other samples and their total anthocyanin contents.

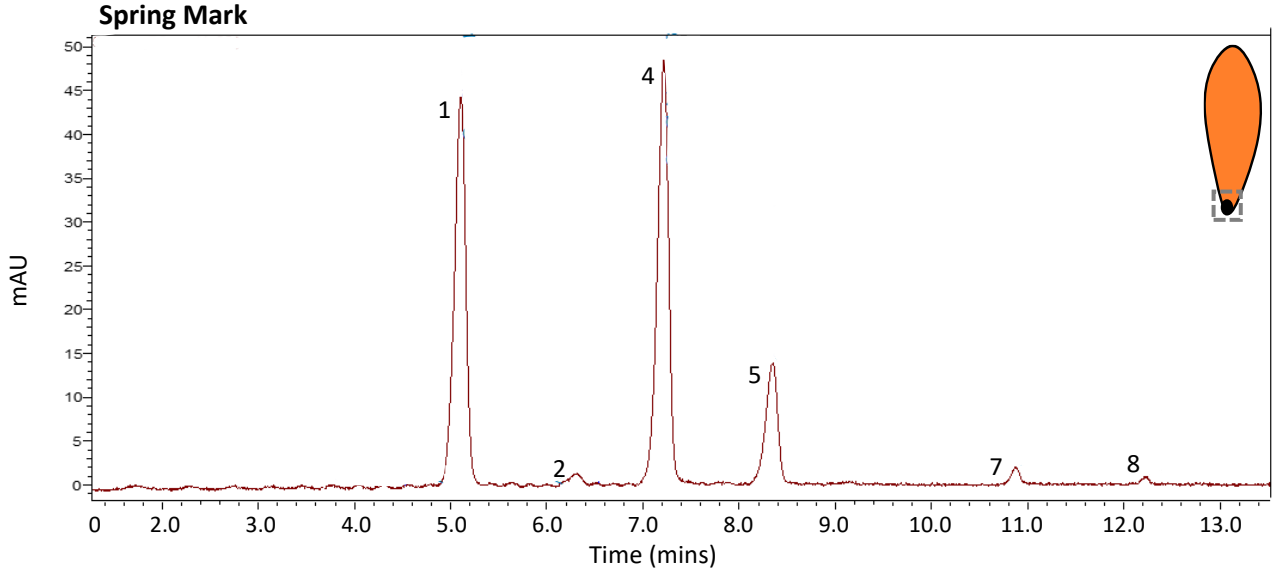
Compound	Retention time (min)	Identity	m/z	HPLC-ESI(+) - MS experiment m/z
1	5.059	CyanGlc	449	MS2 [449]: 287*
2	6.281	CyanGlcMal isomer	535	MS2 [535*]: 287, 401
3	6.484	CyanPentose	535	MS [535]: 240, 287, 331, 403*, 419, 426, 449, 466
4	7.133	CyanGlcMal	535	MS2 [535*]: 287, 449, 491
5	8.283	CyanGlcMeMal	549	MS2 [549*]: 287, 449, 517
6	9.109	CyanGlcCaf	611	MS2 [611]: 231, 258, 287*, 333, 373, 487, 606
7	10.827	CyanGlcCafMal	697	MS2 [697]: 287*, 493, 585
8	12.183	CyanGlCafMeMal	711	MS2 [711]: 287*, 611

Table 4.3. UV spectra and physical properties of all anthocyanins found in the ray floret tissue of *G. diffusa* through HPLC and positive mode electrospray mass spectrometry. Compound number corresponds to peak number in Figure 4.8, * indicates the base peak. Cyan = cyanidin, Glc = glucoside, Caf = cafeate residue, Mal = malonyl residue, MeMal = methylmalonyl residue, Pentose = pentose sugar.

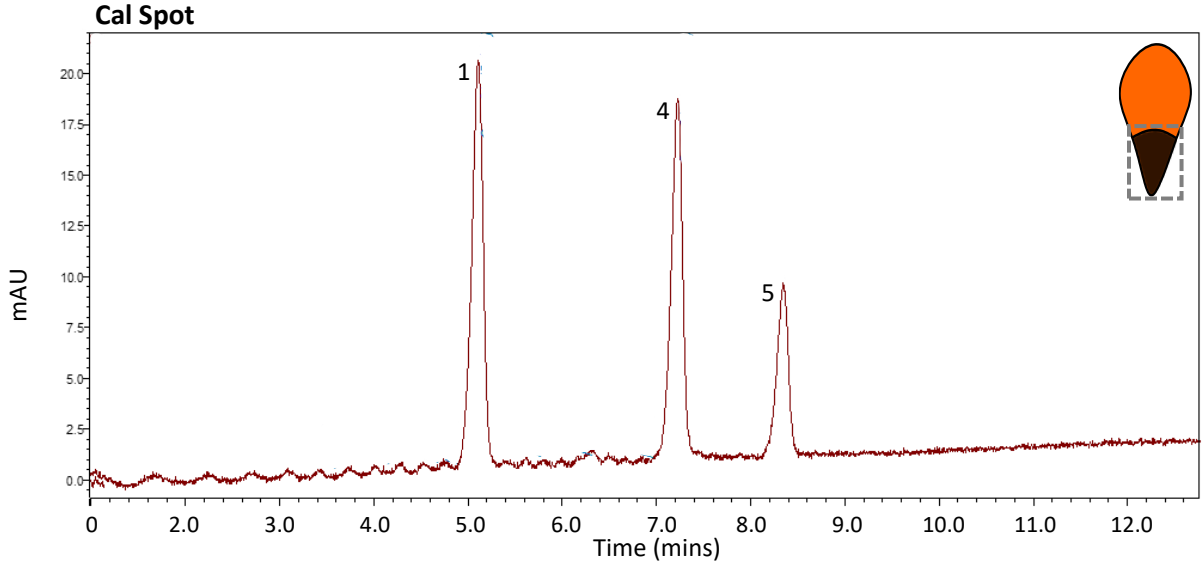
Spring Spot



Spring Mark



Cal Spot



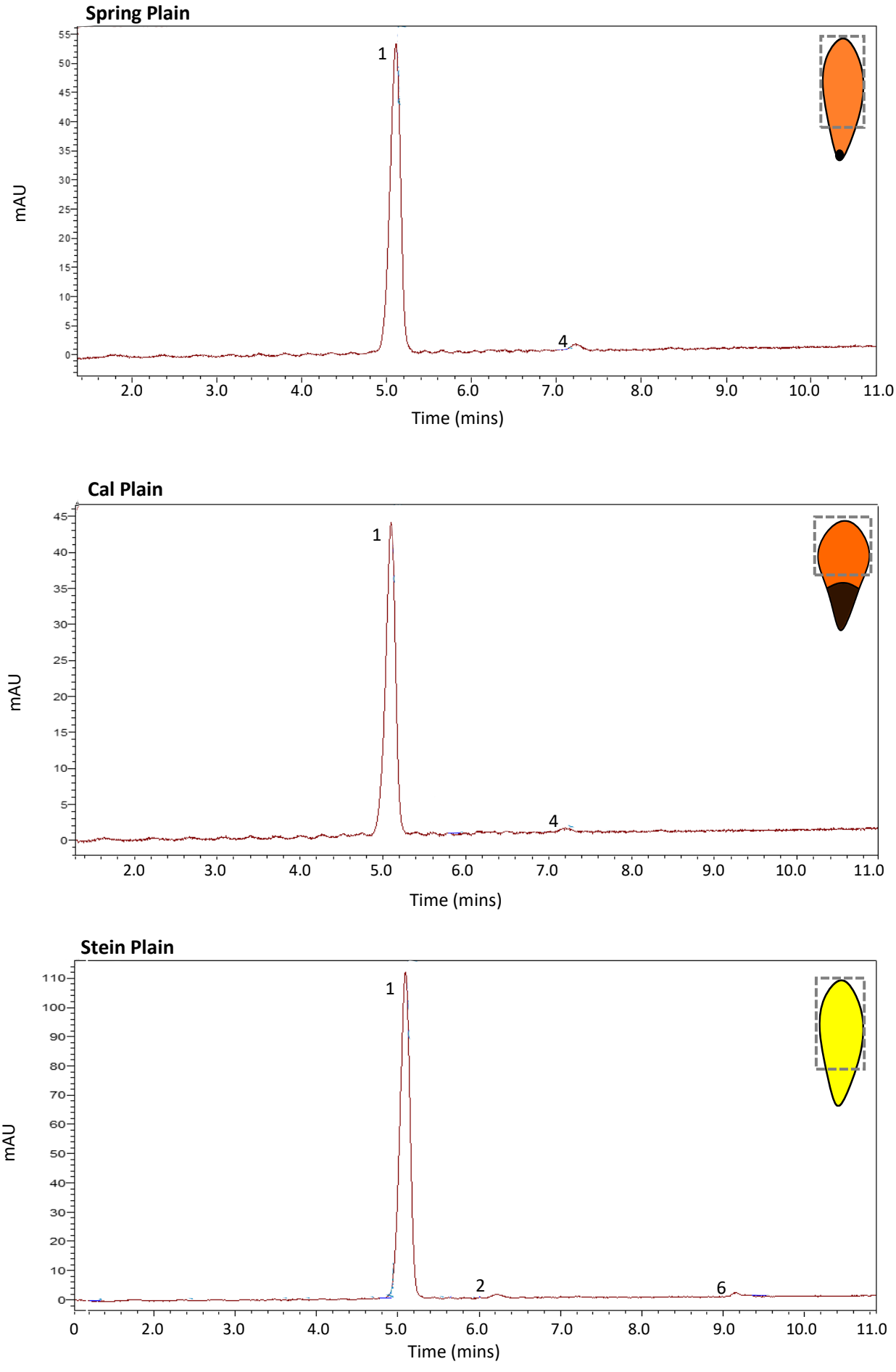


Figure 4.8. High-performance liquid chromatography (HPLC) chromatograms of *G. diffusa* ray floret tissue at 525nm (bandwidth 50nm) – within the absorbance spectra of anthocyanins. Corresponding peaks are numbered throughout, with the MS spectra of each listed in Table 4.3. The tissue sample collected for each analysis is indicated by the grey box on the ray floret diagrams. The sample size for each tissue type was n = 3.

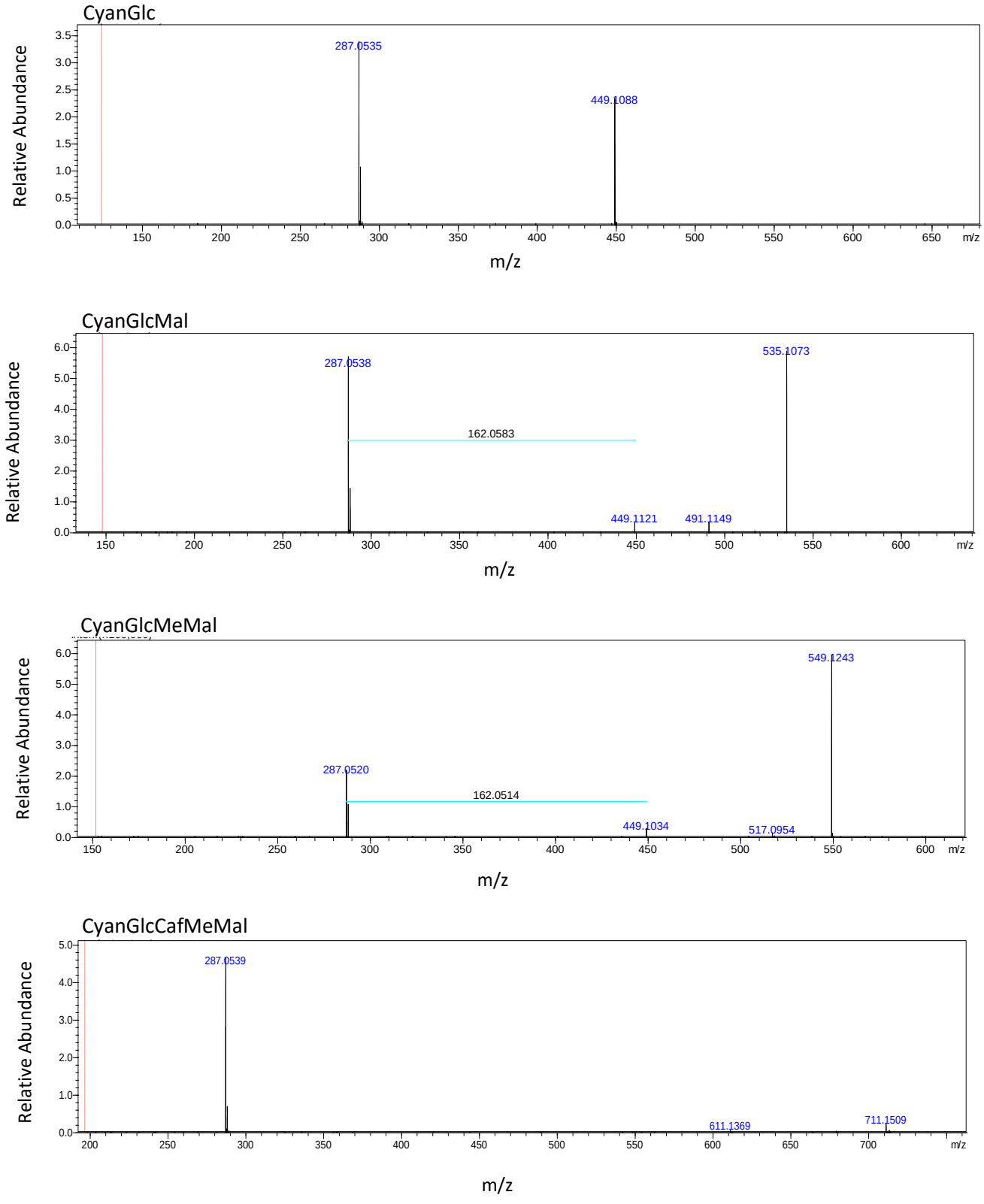


Figure 4.9. Example of mass spectra (MS2) used to identify anthocyanins. The mass spectra shown are from several peaks identified from the *G. diffusa* 'Spring spot' samples. Cyan = cyanidin, Glc = glucoside, Caf = cafeate residue, Mal = malonyl residue, MeMal = methylmalonyl residue. The sample size for each tissue type was n = 3.

4.3.3 The isolation of four homologous *GdMYB8* candidate genes for petal spot pigmentation

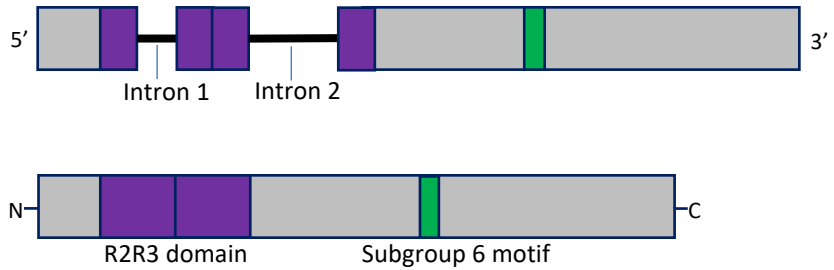
A subgroup 6 R2R3 MYB transcription factor, *GdMYB8a*, was previously identified as a candidate for regulating petal spot pigmentation within the Spring morphotype (Mellers 2016; Walker 2012). *GdMYB8a* was expressed within *G. diffusa* spotted ray florets and capable of producing ectopic anthocyanin production in *Nicotiana tabacum* (Mellers 2016). Three additional homologues were identified through PCR and a transcriptome analysis: *GdMYB8b*, *GdMYB8c*, and *GdMYB8d*. Gene and corresponding amino acid sequences are illustrated in Fig 4.10. These genes were characterised extensively in a single individual and then across multiple individuals. No more than 2 variants of each gene were found per individual, suggesting that these are four recently duplicated genes – rather than alleles of the same gene. This was necessary as no genome sequence is available for *Gorteria* to confirm the number of gene copies. All of the *GdMYB8* genes have a similar structure, with 3 exons and 2 introns (Table 4.4). GdMYB8a - c vary in length by 2 amino acids, while GdMYB8d is 10 - 13 amino acids shorter in length – predominantly due to fewer nucleotides in exon 1. Variation between alleles of the same gene was minimal (Fig 4.10b), with 0 - 4 non-synonymous SNPs identified. Comparing between proteins, GdMYB8a, GdMYB8b, and GdMYB8c had 84 - 91% of amino acids conserved. GdMYB8d is more divergent, with 68 - 69% of its amino acid sequence shared with the other GdMYB8 proteins (for cDNA alignment see Fig 4.11). Examination of the amino acid sequences suggests that the GdMYB8 proteins should all be functional. There are few changes in amino acid sequence within the R2 and R3 MYB DNA binding domains, and the bHLH interaction domain was present in all proteins (Zimmermann et al. 2004). There are some differences in the amino acid sequences that comprise the subgroup 6 motif (in *A. thaliana* KPRPR[S/T]F (Stracke et al. 2001), in *G. diffusa* [Q/K]PQP[S/H][T/K]F). Within GdMYB8a a glutamine (Q) is sometimes found at the first position rather than a lysine (K). GdMYB8d has a histidine (H) at the fifth position - where the other GdMYB8s have a serine (S). The sixth position is a threonine (T) in all GdMYB8s, except GdMYB8c which has a lysine (K). Most amino acid differences between the homologues were found 3' to the subgroup 6 motif, as expected due to the highly conserved N-terminus MYB domains characteristic of R2R3 MYB transcription factors (Stracke et al. 2001).

4.3.4 The conservation of GdMYB8 proteins between morphotypes

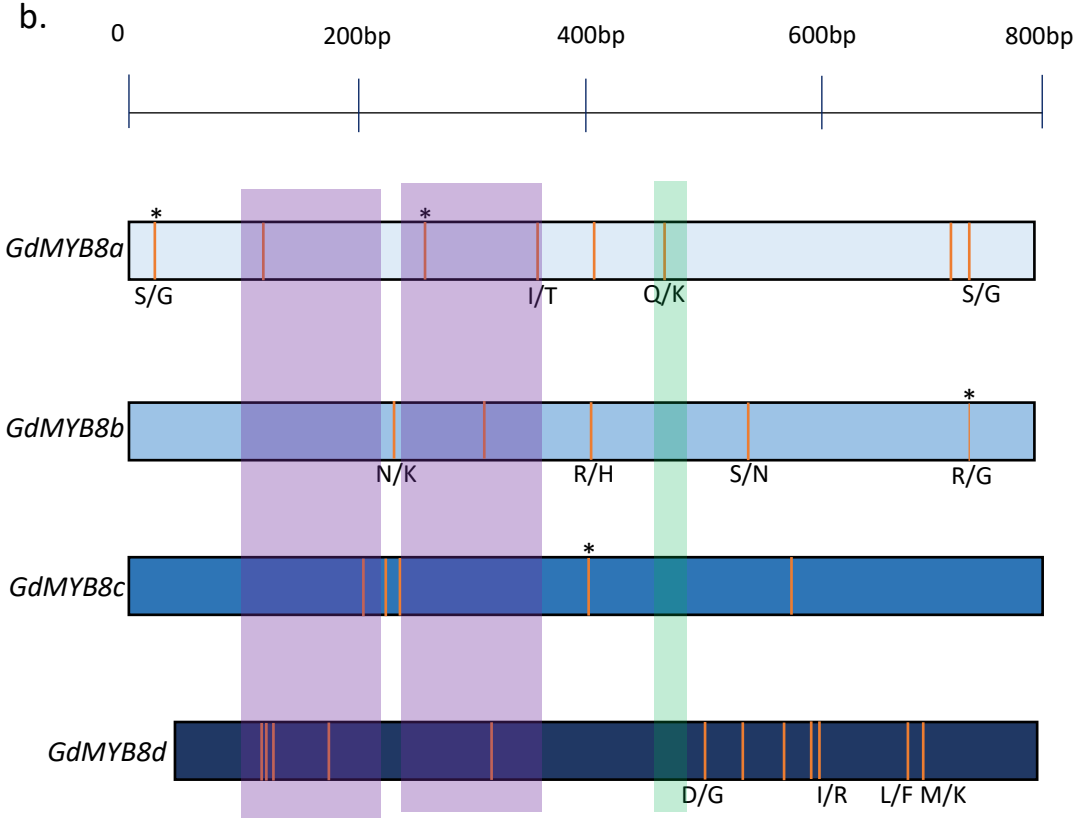
Once *GdMYB8* genes had been characterised in Spring, they were amplified in Cal and Stein. The only gene which was not characterised was *GdMYB8b* in Stein. Attempts using Spring *GdMYB8b* primers were unsuccessful – suggesting SNPs in the primer sequence/s in Stein. An inherited DNA sample mislabelled, due to misidentification of *G. diffusa* morphotype in the field, led to a false characterisation of Stein *GdMYB8b*. The data have therefore been removed, but the mistake was not detected in time for thorough Stein *GdMYB8b* gene hunting to be conducted. Across all three morphotypes the *GdMYB8* genes are highly structurally conserved and very similar in length to the Spring morphotype counterparts. In all 4 genes intron 2 and exon 3 had slightly varying lengths between morphotypes, and in *GdMYB8d* the length of intron 1 also differed between morphotypes (Table 4.4). For GdMYB8a - c, comparisons within a gene found 4 - 8 amino acids were consistently different between morphotypes (Fig. 4.13) (1 - 7% divergence in protein sequence). Most of these differences were in the region 3' of the subgroup 6 motif. Notably, Stein GdMYB8d contained an asparagine (N) where Spring and Cal had a threonine (T) within the subgroup 6 motif, both are amino acids with polar uncharged side chains. Additionally, Cal GdMYB8d had a region with an amino acid deletion followed by 4 amino acids which differed from those found in Stein and Spring. Ultimately, the *GdMYB8* genes appear highly conserved between morphotypes, to conclusively determine

whether the consistent amino acid differences identified are morphotype-specific would require characterising these genes in a greater number of individuals and across the full morphotype geographical range.

a.



b.



c.

Protein 1	Protein 2	Proportion of Identical Amino Acids
MYB8a (264aa)	MYB8b	0.84 - 0.86
	MYB8c	0.85 - 0.87
	MYB8d	0.69
MYB8b (264aa)	MYB8a	0.84 - 0.86
	MYB8c	0.89 - 0.91
	MYB8d	0.69
MYB8c (266aa)	MYB8a	0.85 - 0.87
	MYB8b	0.89 - 0.91
	MYB8d	0.69 - 0.70
MYB8d (254aa)	MYB8a	0.68
	MYB8b	0.68
	MYB8c	0.68 - 0.69

Figure 4.10. Characterising and comparing the subgroup 6 R2R3 *MYB8* genes of the Spring morphotype *G. diffusa*. a) Diagram of the gDNA sequence and amino acid sequence. b) cDNA diagram showing the SNPs found within a gene for each *MYB8*. Each SNP is represented by an orange line. Non-synonymous SNPs have the amino acids listed below, * represent SNPs found to differ between alleles within an individual, the purple and green shading represent the locations of the R2R3 MYB domain and subgroup 6 motif respectively. c) Table demonstrating the proportion of amino acids shared between each MYB8 protein.

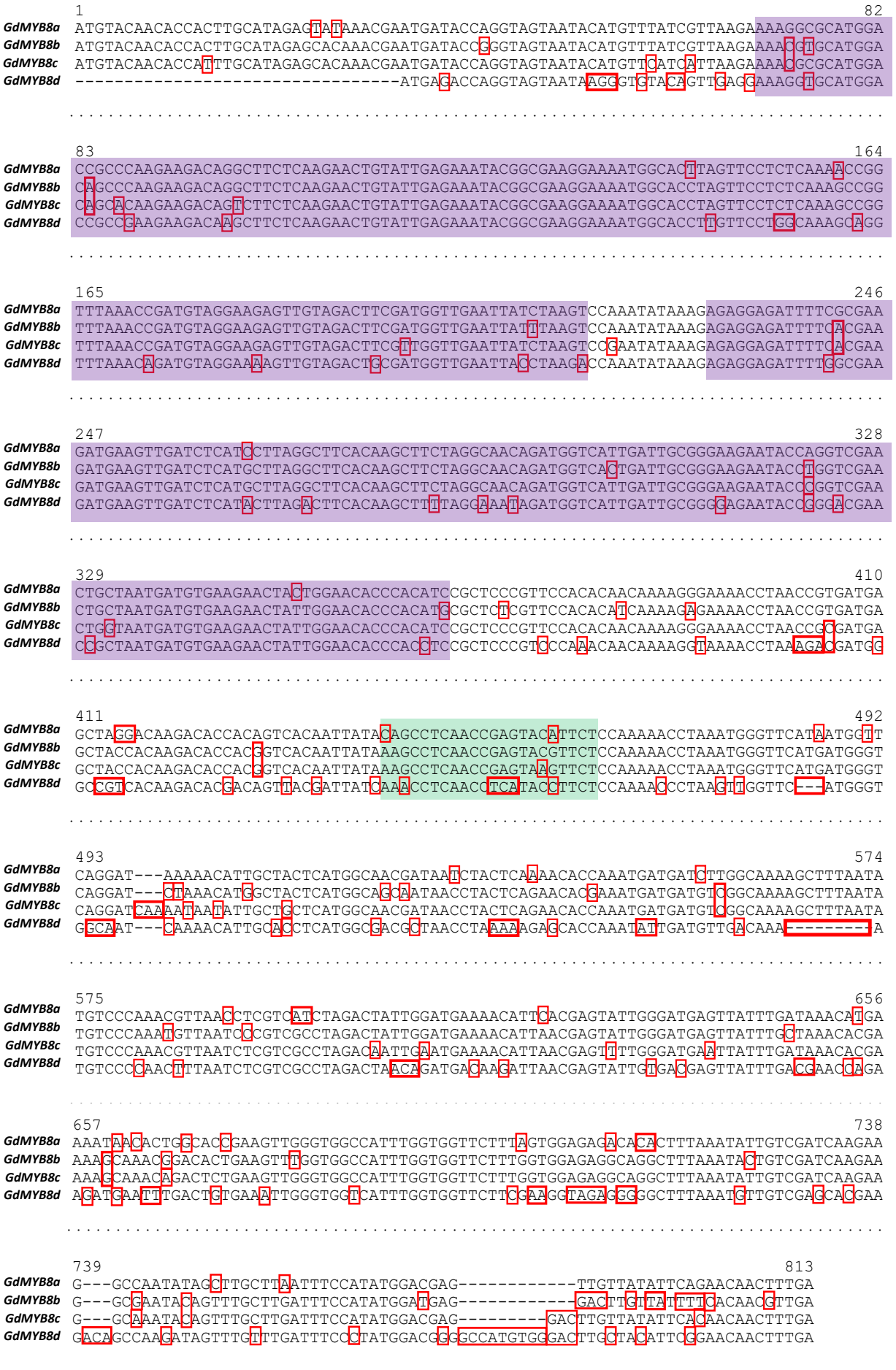


Figure 4.11. MAFFT alignment of the cDNA sequences of each Spring *GdMYB8* gene. SNPs and insertions are indicated by red boxes. The purple and green shading represent the locations of the R2R3 MYB domain and subgroup 6 motif respectively.

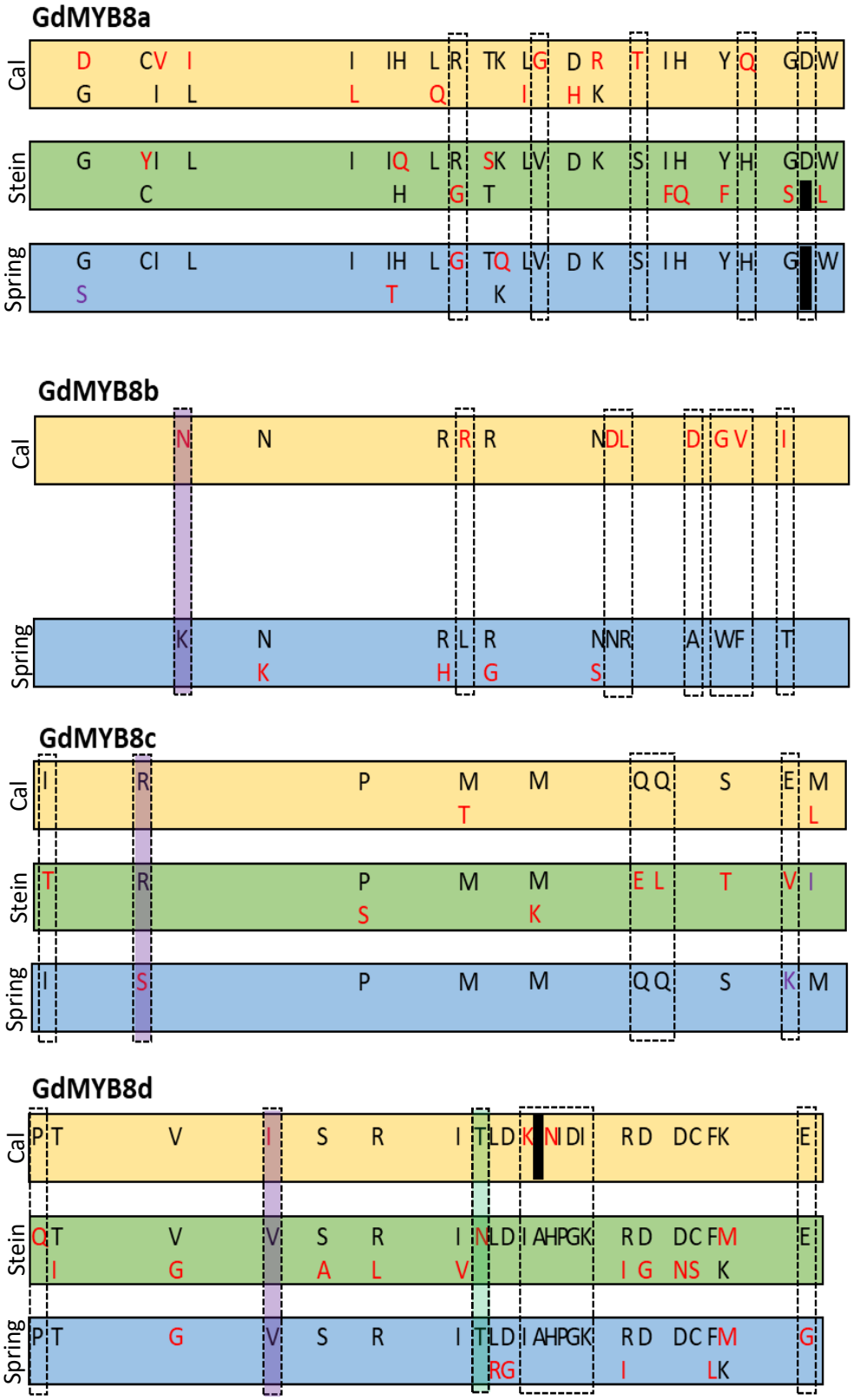
a.

Alignment Position	MYB8a	MYB8b	MYB8c	MYB8d
5	T	T	I	-
9	Y	H	H	-
13	I	I	R	-
15	G/S	G	G	G
20	L	L	S	V
25	G	R	R	G
33	R	R	S	K
54	T	A	A	A
74	N	N/K	N	N
81	R	H	H	G
88	I	M	M	I
111	A	A	G	A
122	I/T	M	I	L
131	G	R	G	V
135	R	R/H	R	D
148	Q/K	K	K	K
153	T	T	K	T
162	I	M	M	M
164	V	G	G	G
167	K	L	Q	-
168	-	-	N	Q
170	I	M	I	I
172	T	T	A	P
175	N	S/N	N	D
176	D	N	D	A
180	K	R	R	R
182	P	R	P	P
186	L	V	V	V
195	T	M	T	T
197	T	I	I	I/R
198	S	P	S	S
200	S	P	P	P
202	L	L	Q	L
204	D	D	N	D
210	Y	Y	F	Y
216	D	A	D	D
220	N	K	K	D
221	N	Q	Q	E
223	G	D	D	D
224	T	T	S	C
227	G	W	G	G
238	H	Q	Q	G
239	T	A	A	A
242	I	T	I	V
247	G/S	G/R	G	D
249	Q	E	K	Q
254	N	D	D	D
260	L	D	D	A
261	-	L	L	M
266	Y	F	Y	H
269	Q	R	Q	Q
208	H	N	N	N
268	E	Q	Q	E
270	L	-	L	L

b.

Alignment Position	MYB8a/b/c	MYB8d
18	T	K
19	C	G
21	S	Q
30	Q	E
52	L	G
72	S	R
126	S	P
127	T	K
134	N	K
137	E	G
138	L	P
139	G	S
152	S	H
157	N	T
159	N	S
160	G	W
163	M	G
165	Q	N
166	D	-
179	L	K
181	T	A
184	D	I
187	G	D/G
189	S	-
190	F	-
191	N	-
194	Q	P
203	L	T
205	E	D
206	N	K
211	W	C
217	K	E
218	H	P
222	T	F
226	V	I
229	P	S
230	F	F/L
234	L	S
235	V	M/K
236	E	V
237	R	E
244	D	E
245	Q	H
248	-	S
250	Y	D
253	L	F
256	H	P
259	E	G
262	-	W
263	-	D
264	-	L

Figure 4.12. The amino acid differences between each Spring MYB8 protein and the position of these changes in an alignment of the complete amino acid sequences. a) Positions where there are amino acid differences between MYB8a - 8c, amino acid differences are highlighted in red, or if there are 3 amino acids at one position the third amino acid is highlighted in green, and blue if there is a fourth amino acid difference. b) Positions that are identical in MYB8a - c but differ in MYB8d. The purple and green shading represent the locations of the R2R3 MYB domain and subgroup 6 motif respectively.



MYB8a	Cal	Stein	Spring
Cal	1 - 0.97		
Stein	0.99 - 0.93	1 - 0.97	
Spring	0.98 - 0.94	1 - 0.95	1 - 0.98

MYB8b	Cal	Stein	Spring
Cal	0		
Stein	n/a	n/a	
Spring	0.97 - 0.95	n/a	1 - 0.98

MYB8c	Cal	Stein	Spring
Cal	1 - 0.99		
Stein	0.98 - 0.96	1 - 0.99	
Spring	0.99 - 0.98	0.98 - 0.97	1

MYB8d	Cal	Stein	Spring
Cal	1		
Stein	0.97 - 0.93	1 - 0.96	
Spring	0.97 - 0.95	0.99 - 0.94	1 - 0.98

Figure 4.13. Comparison of amino acids within GdMYB8 proteins between the *G. diffusa* morphotypes Cal (yellow), Stein (green), and Spring (blue). The black dashed boxes indicate sites where the amino acid found is consistently different between morphotypes – boxes shaded purple are in the R2R3 MYB domain, and those shaded green are in the subgroup 6 motif. Black shading indicates a gap of one amino acid in the alignment. For each protein, matrices give the proportion of amino acids shared between and within morphotypes.

Gene	Morphotype	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	gDNA	cDNA	Protein
MYB8a	**Cal**	162	126	129	476	507	1400	798	265
	Stein	162	126	129	472	507	1396	798	265
	Spring	162	126	129	474	504	1395	795	264
MYB8b	**Cal**	162	126	129	457	504	1378	795	264
	Spring	162	126	129	455	504	1376	795	264
MYB8c	**Cal**	162	130	129	476	510	1407	801	266
	Stein	162	130	129	487	510	1418	801	266
	Spring	162	130	129	476	510	1407	801	266
MYB8d	**Cal**	127	113	129	216	506	1091	762	253
	Stein	127	128	129	219	509	1112	765	254
	Spring	127	109	129	226	509	1100	765	254

Table 4.4. The composition of each *GdMYB8* gene within Cal, Stein, and Spring. The number of base pairs in each exon and intron is given. The length of the full genomic DNA and complementary DNA is given, along with the number of amino acids in the protein.

4.3.5 *Gorteria* MYB8 homologues cluster within the Asteraceae subgroup 6 R2R3 MYB transcription factor clade

All available Asteraceae subgroup 6 R2R3 MYB transcription factors were used to construct a maximum likelihood amino acid tree (Fig 4.14). *Gorteria* MYB8 amino acid sequences cluster within the subgroup 6 R2R3 MYB clade, with reasonably high bootstrap support. MYB8d is basal to the clade, and MYB8b and MYB8c are sister to one another. *Gorteria personata* MYB8 sequences cluster with the corresponding GdMYB8 protein (i.e. GpMYB8a is in a clade with GdMYB8a), demonstrating that the gene duplication events did not occur within *G. diffusa*. The absence of *G. personata* MYB8d does not indicate its absence from this species, as thorough gene hunting was not conducted. The clustering of *Gorteria* MYB8 proteins demonstrates that the duplication event producing this gene family was fairly recent. Higher taxonomic sampling resolution would enable a less ambiguous conclusion as to the point within Asteraceae evolution at which these duplication events occurred. Of the Asteraceae sequences available, the only other representative of the subfamily Cichorioideae (containing around 3000 species) was *Lactuca sativa* (Panero and Funk 2008).

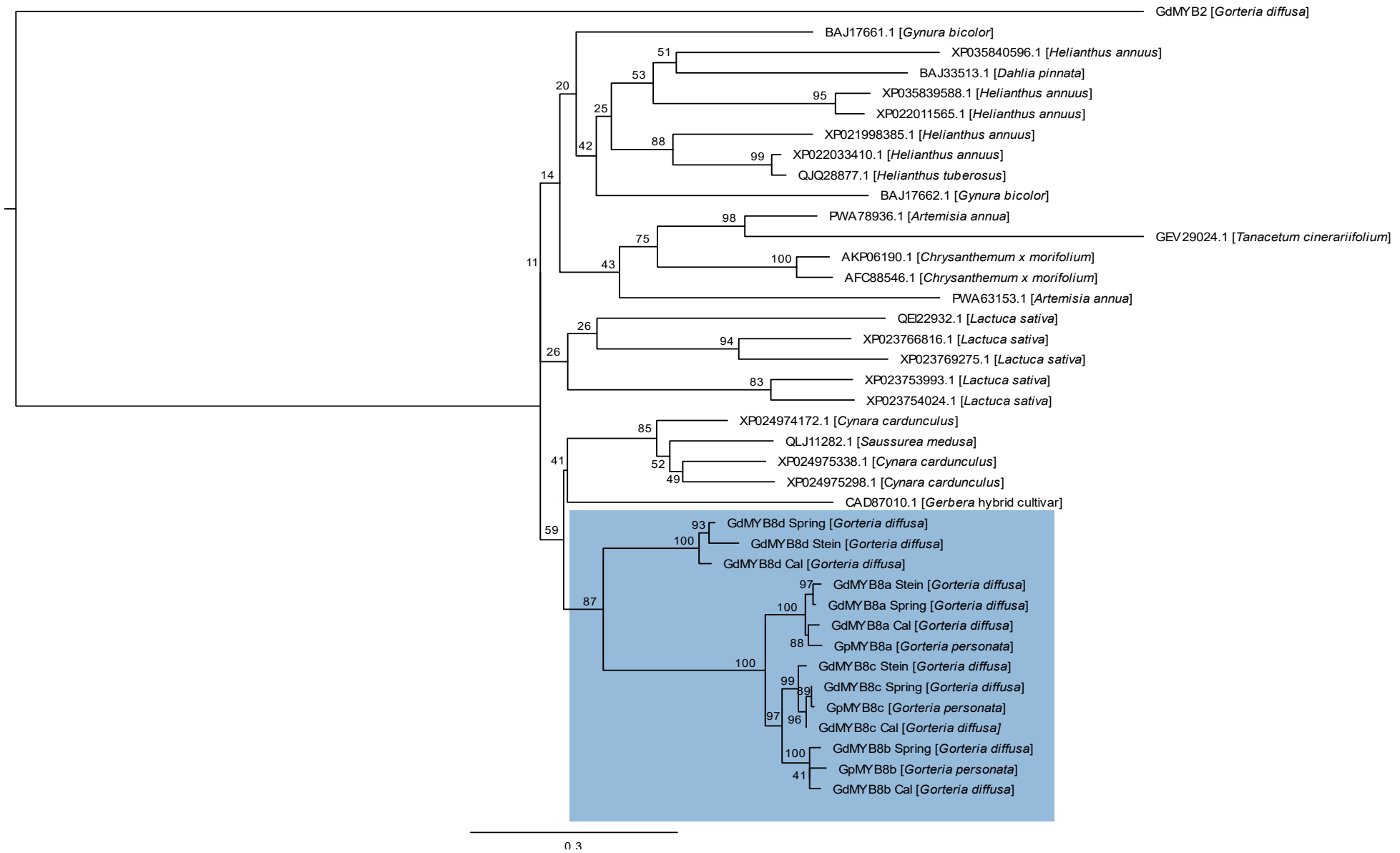


Figure 4.14. Maximum likelihood phylogeny of Asteraceae subgroup 6 R2R3 MYB amino acid sequences. This phylogeny was constructed from a RAxML analysis, using 200 bootstrap resamplings. Bootstrap values are indicated. The tree was rooted with GdMYB2, a *G. diffusa* subgroup 9 R2R3 MYB (Murphy 2009). *Gorteria* MYB8 genes sequenced during this project are indicated by the blue box.

4.3.6 Three *GdMYB8* genes are upregulated in petal spots during development

Note. The following abbreviations are used: Cal spotted (Sp) ray floret petal segments and plain top (Tp) segments of ray floret petals. Spring whole spotted ray (Sr) floret petals and whole plain ray floret (Pr) petals. Developmental stage 1 (1) and 2 (2) Diagrams of petal sections are in Fig 4.15 and Fig 4.16.

The expression patterns of each *GdMYB8* gene were compared between spotted and plain ray floret petal tissue at two developmental stages during spot development. This comparative analysis was conducted in Cal and Spring (Fig 4.15), while plain petal tissue was also analysed in Stein (Fig 4.16). In both Cal and Spring morphotypes *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* were significantly upregulated in spotted tissue compared to plain tissue at both developmental stages (results of significance tests in Table 4.5). The expression levels of *GdMYB8a* and *GdMYB8c* in Stein were very low, consistent with the Spring and Cal results indicating that these genes are not expressed or have extremely low expression in plain petal tissue. In both Cal and Spring, within-morphotype comparisons showed there was no difference in expression of *GdMYB8d* between spotted and plain petal tissues. During the second developmental stage, *GdMYB8d* was significantly upregulated in spotted and plain petal tissues of Cal (Sp1 – Sp2: $t = -10.92$, $p < 0.0001$; Tp1 – Tp2: $t = -6.84$, $p = 0.0002$) and Stein tissue (Pr1 – Pr2: $t = -2.94$, $p = 0.042$). Spring had similar *GdMYB8d* expression patterns to Stein and Cal (Fig 4.15, Fig 4.16), but *GdMYB8d* upregulation in Spring between developmental stages was not statistically significant. In both Cal and Spring, *GdMYB8b* expression at developmental stage 2 was significantly higher than any other gene across all tissue types (data not shown). *GdMYB8b* expression levels differed significantly in the spotted tissue at different developmental stages (Sp1 - Sp2: Cal $t = -4.24$, $p = 0.006$; Sr1 – Sr2: Spring $t = -7.64$, $p = 0.0003$). *GdMYB8a* expression also differed in spotted tissue between developmental stages in Spring ($t = 5.00$ $p = 0.005$). In Cal spotted tissue *GdMYB8b* also had significantly higher expression than the other genes during developmental stage 1 (*MYB8a* Sp1 – *MYB8b* Sp1 $t = -5.66$, $p < 0.0001$; *MYB8b* Sp1 – *MYB8c* Sp1 $t = 4.256$ $p = 0.0005$). In Spring, *GdMYB8c* had significantly higher expression at developmental stage 1 than *GdMYB8b* (Sr1 $t = -2.49$ $p = 0.027$). In both Spring and Cal, *GdMYB8c* expression in spotted petal tissue was higher than that of *GdMYB8a* at equivalent developmental stages, with the exception of Cal developmental stage 1 (Cal Sp2 $t = -3.16$ $p = 0.006$; Spring Sr1 $t = -2.97$ $p = 0.0093$, Sr2 $t = -3.05$ $p = 0.008$).

The relative expression of genes was generally much higher in Cal than in Spring. However, the data are not directly comparable because different tissue segments were used, some of the qRT-PCR primers differed (due to SNPs between morphotypes within a gene), and complete equivalence in development stage cannot be assured. Trends in the expression level can still be qualitatively compared. It is interesting to note that while in Cal spotted tissue *GdMYB8c* expression appears to increase from developmental stage 1 to stage 2, in Spring expression of this gene decreases between developmental stages, but this requires further elucidation. Comparing expression levels within a morphotype, the relative difference in expression between *GdMY8b* and *GdMYB8c* is greater in Cal than in Spring.

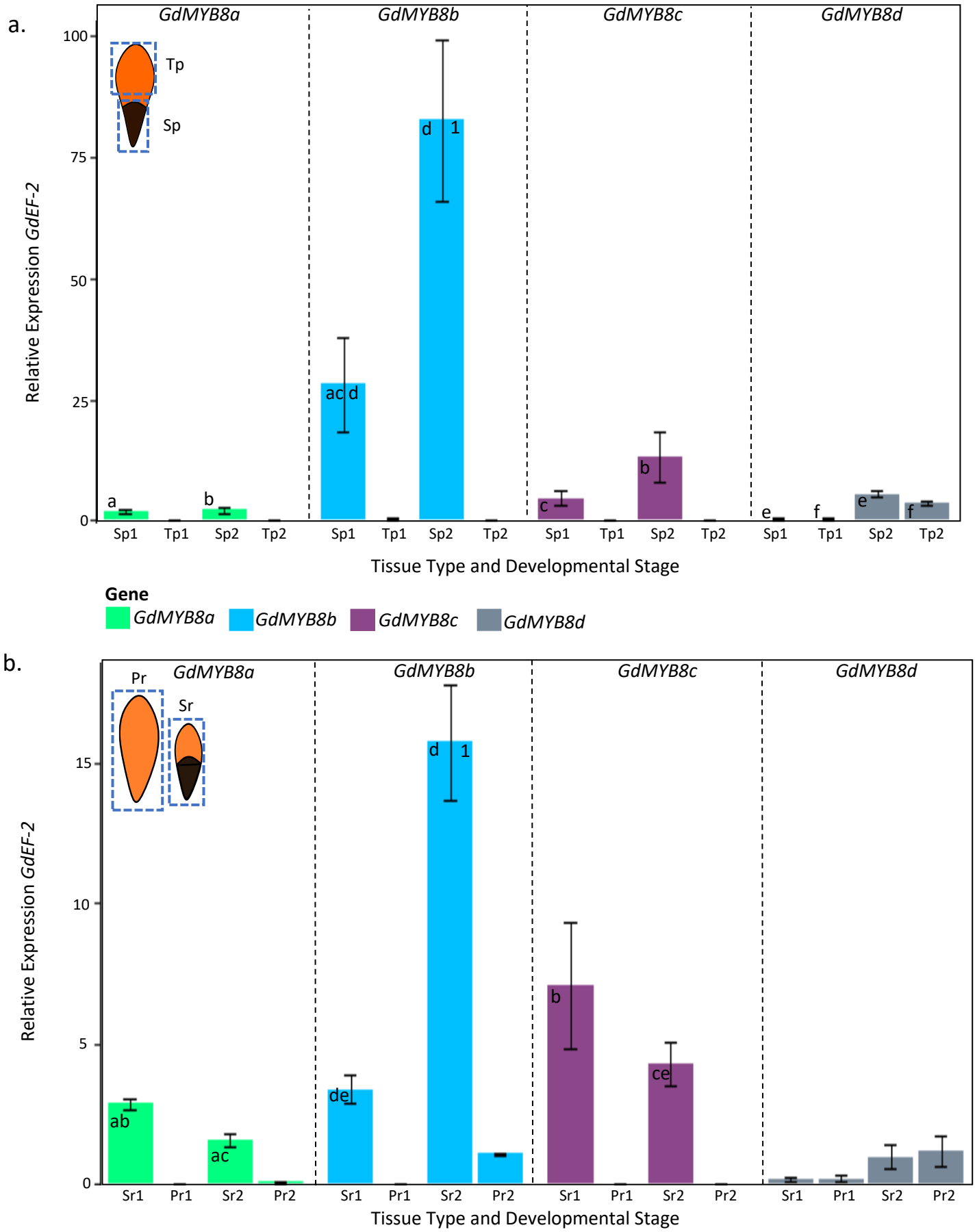


Figure 4.15. qRT-PCR results showing the relative expression of each *GdMYB8* gene at two developmental stages in a) Cal spotted (Sp) and plain (Tp) ray floret petal tissue b) Spring whole spotted ray florets (Sr) and whole plain ray florets (Pr). The tissue segments used are indicated in the ray floret diagrams for each morphotype. Error bars represent mean ± s.e. The sample size for each tissue type was n = 3. Bars that share letters had significantly different expression levels (determined by a t-test, $p \leq 0.05$) within a morphotype. "1" indicates that spotted tissue at developmental stage 2 had expression levels significantly higher that all other spotted tissue types at both developmental stages. For *GdMYB8a* - *c* spotted tissue had significantly higher expression levels than non-spotted tissue at both developmental stages. Only biologically relevant comparisons of expression levels were analysed through statistical tests.

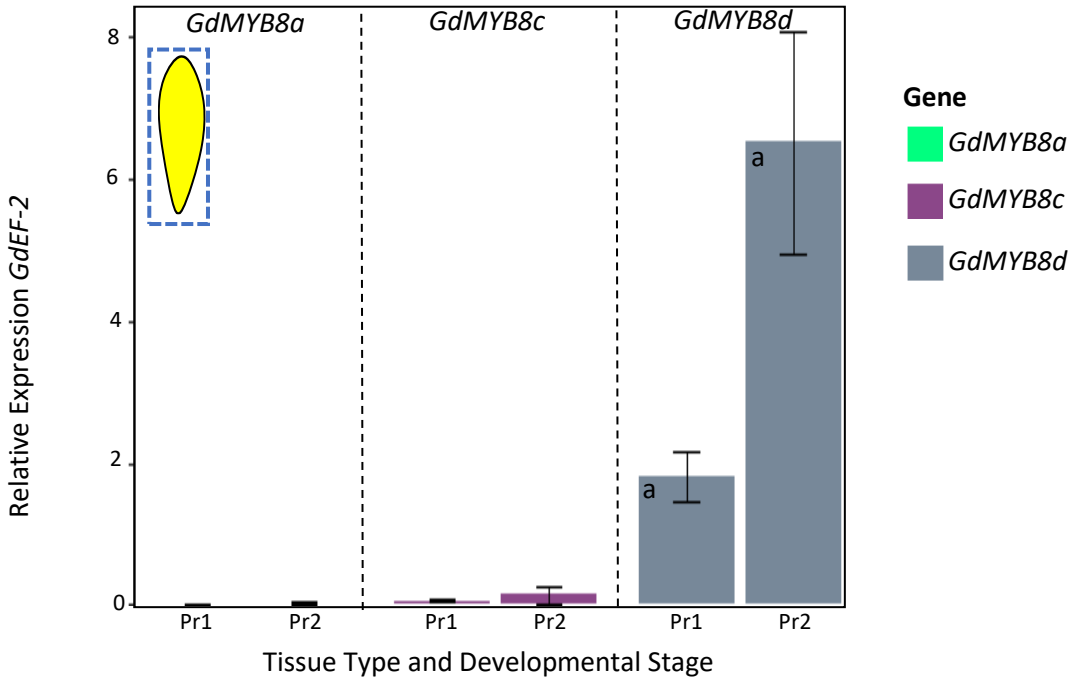


Figure 4.16. qRT-PCR results showing the relative expression of each *GdMYB8* gene at two developmental stages in Stein. Only non-spotted Stein capitula were used in the analysis, so all samples are plain ray florets (Pr). *MYB8b* is yet to be characterised in Stein. The tissue segments used are indicated in the ray floret diagrams for each morphotype. Error bars represent mean ± s.e. The sample size for each tissue type was n = 3. Bars that share letters had significantly different expression levels (determined by a t-test, $p \leq 0.05$) within a morphotype. *GdMYB8d* had significantly higher expression levels than *GdMYB8a* and *GdMYB8c* at both developmental stages.

Morphotype	Gene	Samples	t.ratio	p.value
Cal	*MYB8a*	Sp1 - Tp1	7.54	0.0003*
		Sp2 - Tp2	7.9	0.0002*
	MYB8b	Sp1 - Tp1	6.35	0.001*
		Sp2 - Tp2	10.64	<.0001*
	MYB8c	Sp1 - Tp1	5.56	0.0024*
		Sp2 - Tp2	7.95	0.0002*
	MYB8d	Sp1 – Tp1	0.060	0.999
		Sp2 - Tp2	1.87	0.311
Spring	*MYB8a*	Sr1 - Pr1	-19.29	<.0001*
		Sr2 - Pr2	-12.4	<.0001*
	MYB8b	Sr1 - Pr1	-13.91	<.0001*
		Sr2 - Pr2	-11.92	<.0001*
	MYB8c	Sr1 - Pr1	-8.06	0.0002*
		Sr2 - Pr2	-6.44	0.0009*
	MYB8d	Sr1 – Pr1	-0.303	0.990
		Sr2 - Pr2	0.236	0.995

Table 4.5. Significance values of pairwise comparisons from the qRT-PCR expression data, highlighting results that indicate whether or not each gene is upregulated within the petal spot during development. Spotted and plain ray floret petal tissue are compared within a gene at the same developmental stage for *GdMYB8a – GdMYB8c*. Student's t-tests were used. Significant results ($p < 0.05$) are indicated with an asterisk.

4.4 Discussion

This research has established that *G. diffusa* ray floret petals are pigmented by cyanidin glucosides. Cyanidins containing a malonyl group have a major role in pigmenting petal spots but not plain petal regions. Four MYB transcription factors potentially involved in regulating the production of anthocyanin in the ray floret petals were identified as subgroup 6 R2R3 MYBs. The phylogenetic position and sequence homology of these *GdMYB8* genes suggests they are recent duplicates. Three of the genes were upregulated during petal spot formation in the developing spotted ray florets, confirming that they are good candidates for regulation of spot-specific anthocyanin production. Anthocyanin composition, protein sequences, and gene expression patterns were generally consistent across different floral morphotypes of *G. diffusa*, although subtle differences were identified.

Spring complex petal spots had significantly higher anthocyanin content than all other tissues across the morphotypes analysed, this included the complex petal spots of Cal that appear to be heavily pigmented. Spring petal spots contain swollen epidermal cells forming papillae that are filled with anthocyanin (Thomas et al. 2009); the presence of these cells, that are absent in Cal, could potentially explain the differential anthocyanin content between the spots of the two morphotypes. Cal petal spots did not contain more anthocyanin than plain petal regions. The raised appearance of Cal spots is due in part to curvature of the ray floret itself rather than enlarged pigmented epidermal cells. Pigment on the abaxial side of the Cal petals is rarely present at the curved basal region of the ray floret petals and, as such, the anthocyanin content of the spotted region is not a result of cumulative anthocyanin content from the spot and abaxial surface but is likely solely from the spot. It is possible that other forms of pigmentation are also important in forming the dark green of the Cal petal spot, and anthocyanin is not as dominant as in the Spring morphotype. Supporting this, Walker (2012) found significantly higher chlorophyll content in Cal spots than in Spring spots. All plants sampled were grown in the same environmental conditions, but the quantity of abaxial pigmentation may be influenced by environmental factors – explaining the spectrum of colouration (from light to dark purple) seen in the plain ray floret petals of Spring sampled in the field. In *Petunia*, anthocyanin accumulation on abaxial petal surfaces of the bud is light-regulated (Albert et al. 2011). A controlled experiment growing *G. diffusa* plants in different light-conditions could provide a better understanding of the nature of abaxial pigmentation.

Analyses of anthocyanin composition found that in all cases, irrespective of morphotype or spot phenotype, anthocyanin distribution in the petals followed similar patterns. The predominant anthocyanin across all samples was cyanidin 3-glucoside, and all anthocyanins detected were derived from cyanidin (with one possible unconfirmed exception). In flowers of *Chrysanthemum morifolium* and *Lilium* spp. cyanidin derivates contribute toward pink to red colouration in flowers (Hong et al. 2015; Suzuki et al. 2016), whereas in cornflower and *Meconopsis grandis* they facilitate a blue floral colouration (Yoshida et al. 2006; Yoshida and Negishi 2013). A large proportion of the anthocyanins pigmenting petal spots were cyanidins acylated by malonate, while cyanidin with malonyl groups were absent or present in very small quantities across plain petal tissues of all morphotypes. Malonic acid is the most frequent aliphatic acyl group in acylated anthocyanins. Malonylated anthocyanins are found throughout the Asteraceae family, including in *Senecio cruentus* and *Gerbera* (Harborne 1963; Takeda et al. 1986). Takeda et al. (1986) speculated that malonic acid may be the most usual acyl substituent in the Asteraceae, based on their findings that anthocyanins containing a malonyl group occurred in representatives from five different tribes. Anthocyanins acylated by malonic acid can form anthocyanin zwitterions in the vacuolar sap, where protons disassociate and decrease the pH value of

the solution (Takeda et al. 1986). The increased acidity of the medium protects the anthocyanin from degradation induced by pH increases. As such, malonic acid acylation can increase the stability of anthocyanins (Figueiredo et al. 1999). In *C. gracilis* differential pigmentation was also found between spotted and plain petal tissues, although in this case the discrepancy was in the anthocyanidins present (cyanidin and peonidin occurred only within spots) (Martins et al. 2013), rather than the acylation of the anthocyanins as we demonstrate here in *G. diffusa*.

A more subtle difference in anthocyanin composition was found between morphotypes, with all Cal tissues lacking anthocyanins containing caffeate. Anthocyanins with caffeate were present in Spring complex spots and simple spots ('marks'), and Stein plain petal tissues in relatively small quantities. Intraspecific differences in anthocyanin composition, regarding glucose and malonate residues, have been reported between cultivars within other species belonging to the Asteraceae family; these include *Dahlia* and *Gerbera jamesonii* (Takeda et al. 1986). In the latter, the cultivars differing in anthocyanins had very similar floral colouration despite these differences, a phenomenon also demonstrated in poinsettia bracts (Stewart et al. 1979, 1980), geranium florets (Asen 1983) and lily tepals (Nørbæk and Kondo 1999). Evidently, it is difficult to conclude how or if these differences in anthocyanin composition contribute toward divergent spot phenotypes within *G. diffusa*. It is interesting to note that the simple spots at the base of the plain Spring ray florets have approximately the same anthocyanins present as the complex spots of Spring that have much more elaborate colouration.

A previously conducted phylogenetic analysis demonstrated that *GdMYB8a* is well supported within the subgroup 6 R2R3 MYB clade (Mellers 2016). The phylogeny produced here considers the placement of all GdMYB8 homologues within a phylogeny of Asteraceae subgroup 6 R2R3 MYBs. *MYB8a-c* were also amplified in *G. personata*, and the predicted amino acid sequences of these genes clustered with the orthologous GdMYB8 amino acid sequences. This demonstrates that the duplication events from which MYB8a, MYB8b, and MYB8c derive, occurred prior to the divergence of *G. diffusa* and *G. personata*. From examination of the *GdMYB8* amino acid sequences, it appears that the proteins should be functional. Protein length is similar to that seen in other subgroup 6 R2R3 MYBs and there are no premature stop codons or large indels that could alter protein structure. Comparing between GdMYB8 proteins there are minimal amino acid differences within the conserved R2 and R3 MYB domains, and the regions encoding the bHLH binding domain and subgroup 6 motif are present in all *GdMYB8* genes. There are a few amino acid differences between GdMYB8 proteins within the subgroup 6 motif, but it is unclear as to whether this has implications for protein functioning as this motif has not yet been linked to molecular functions (Millard et al. 2019). Consistent with other R2R3 MYB proteins, the majority of sequence divergence between genes occurred 3' of the subgroup 6 motif. There is little structural information on this section of the protein from other species. Generally, outside of DNA binding domain regions, plant transcription factors are predicted to have extensive disordered regions that enable dynamic interactions with other partners. Recently, Millard et al. (2019) identified a bHLH interaction motif within the non-conserved region of an *A. thaliana* R2R3 MYB, demonstrating a correlation between the affinity of the MYB - bHLH interaction and the phenotypic output controlled by the MBW complex. This indicates that divergence in these regions could play a role in functional specialisation. However, many of the MYBs that regulate anthocyanin biosynthesis have considerable sequence divergence outside of the MYB domains and identical functions even between species, for example, petunia AN2 and maize C1 are functionally interchangeable (Quattrocchio et al. 1999; Ramsay and Glover 2005). In *G. diffusa* there are relatively

few amino acid differences between genes, particularly *GdMYB8a* – *GdMYB8c,* so it is likely that protein structure is extremely similar. Nevertheless, these considerations highlight the importance of conducting functional analyses.

Of the four candidate *GdMYB8* genes examined for regulating anthocyanin synthesis within the petal spots of *G. diffusa*, three showed significant differential expression between spotted and plain petal regions. This upregulation in spotted tissue was evident in *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* at both developmental stages tested, just after spot initiation and as specialised cell types begin to form. The level of expression in spotted tissue differed between the three genes. *GdMYB8b* had very high expression levels relative to the other genes at the second developmental stage, in both Cal and Spring spotted tissue. In Cal, *GdMYB8b* also had the highest expression level at the first developmental stage, whereas in Spring *GdMYB8c* was expressed at higher levels. This suggests that *GdMYB8b* and *GdMYB8c* are the dominant genes in regulating petal spot pigmentation. In both morphotypes there was a large difference in the expression levels of *GdMYB8b* between spot developmental stages. In the Spring morphotype this is the stage during which the papillae cells fill with anthocyanin (Thomas et al. 2009). As Cal ray floret petals were split latitudinally into spotted and plain sections, while whole spotted and plain ray florets were used in Spring, we cannot compare expression levels quantitatively between the two morphotypes. Additionally, there is only rough equivalence of developmental stages between the morphotypes.

GdMYB8d did not show differential expression between spotted and plain petal tissue but was upregulated in both petal regions at the second developmental stage. As such, *GdMYB8d* is a good candidate for regulating abaxial pigmentation that first becomes visible during the second developmental stage. However, *GdMYB8d* is also upregulated within Spring spotted ray florets at the second developmental stage and these ray florets have no visible abaxial pigmentation. As such, this gene could also be involved in regulating anthocyanin synthesis within the spot at a later developmental stage. The expression patterns of the *GdMYB8* genes in non-spotted Stein individuals confirmed the patterns inferred from Spring and Cal, with *GdMYB8a-c* having little to no expression within these plain ray floret petals and *GdMYB8d* expressed at both developmental stages, with significant upregulation at developmental stage 2 coinciding with the development of abaxial pigmentation. There could also be anthocyanins contributing to the deep orange colouration of adaxial plain regions, thought to predominantly be comprised of carotenoid pigmentation. *In situ* hybridisation could be used in future to determine the spatial location of *GdMYB8d* expression and test its appropriateness as a candidate for regulation of abaxial pigmentation.

Overall, the gene expression patterns suggest that sub/neofunctionalization may have occurred between *GdMYB8a-c* and *GdMYB8d*, as the latter shows highly divergent expression patterns compared to the other *GdMYB8* genes. *GdMYB8a, GdMYB8b*, and *GdMYB8c* are good candidates for petal spot anthocyanin regulation. Interestingly, the expression patterns were quantitatively different between these genes. As such, GdMYB8a, GdMYB8b, and GdMYB8c could have divergent roles in spot pigmentation or, alternatively, the gene duplication may have been recent enough that functional differentiation or degeneration of redundant copies has not yet occurred.

Chapter 5. The potential downstream targets of GdMYB8

Note. Enzyme names are abbreviated in the main text but names are written in full in Table 5.1.

5.1 Introduction

The anthocyanin pathway is well characterised in terms of biochemistry, distribution, biosynthesis and regulation (Grotewold 2006; Winkel-Shirley 2001). The genes encoding the structural enzymes of the pathway have been identified in many species. The first committed enzymatic reaction of the flavonoid pathway is catalysed by chalcone synthase (CHS), which synthesises tetrahydroxy-chalcone from three molecules of malonyl-CoA and one molecule of *p*-coumaroyl-CoA (pathway outlined in Fig 1.1). The common precursor of anthocyanins, dihydroflavonols, are produced by a series of sequential reactions and early biosynthetic genes (EBGs) (*CHS*, *CHI*, *F3H*, and *F3'H*) encode the required enzymes. Downstream enzymes are encoded by late biosynthetic genes (LBGs) (*DFR*, *ANS*, and *UFGT*) (Grotewold 2006). Dihydroflavonol 4-reductase (DFR) and its cofactor (NAPDH) catalyse the first committed step for anthocyanin (and proanthocyanidin) synthesis, producing leucoanthocyanidins (Petrussa et al. 2013; Shi and Xie 2014). DFR competes with FLS for dihydroflavonols, of which there are three types. In some plant species DFR can use any one of the dihydroflavonol substrates, while in others it is more restricted. Anthocyanidin synthase (ANS) is an oxygenase enzyme that catalyses the subsequent enzymatic reaction, oxidising leucoanthocyanidins into the corresponding anthocyanidins (pelargonidin, cyanidin, or delphinidin) (Deng and Lu 2017). These anthocyanidins are then modified, often in a taxon-specific manner, through glycosylation, acylation, and methylation. Typically, UDP-3-O-glucosyltransferase (UFGT) enzymes glycosylate anthocyanins through the addition of a glucose moiety, forming the stable end product anthocyanin. Anthocyanins are acylated by anthocyanin acyltranferases, which are a class of BAHD family proteins that transfer an acyl-CoA acyl group to an anthocyanin sugar moiety. Malonyl transferases, for example, catalyse the addition of a malonyl group (Nakayama et al. 2003). Acylation contributes toward the stability of the anthocyanin and can influence pigment colouration (Springob et al. 2003). Distinct branches of the flavonoid pathway compete for common substrates, for example, in *A. thaliana* seeds the anthocyanin content increases with the loss of proanthocyanidin synthesis (which requires dihydroflavonols) and reduction of anthocyanidin reductase activity (Albert et al. 1997; Xie et al. 2003). As such, a combination of factors within the cell determine which class of pigments are synthesised (Liu et al. 2017).

Subgroup 6 and subgroup 7 R2R3 MYB transcription factors are associated with activation of anthocyanin synthesis across a diverse range of taxa (Albert et al. 2011; Carey et al. 2004; Cone et al. 1986; Gonzalez et al. 2008; Goodrich et al. 1992; Lin-Wang et al. 2010; Ludwig et al. 1989; Paz-Ares et al. 1986, 1987; Quattrocchio et al. 1999; Schwinn et al. 2006; Spelt et al. 2000). A single transcription factor often regulates several steps within the anthocyanin pathway. The majority of eudicot species investigated have MYBs that regulate EBGs and different MYBs regulating LBGs, although partial overlap in regulation sometimes occurs (Petroni and Tonelli 2011). The enzymes that fall into each regulatory cluster vary across taxa (e.g. Martin et al. 1991; Pelletier and Winkel-Shirley, 1996; Quattrocchio et al. 1993). In *A. thaliana*, EBGs of the anthocyanin pathway (*AtFLS, AtCHS, AtCHI, AtF3H* and *AtF3'H*) are regulated by subgroup 7 R2R3 MYB proteins (AtMYB11, AtMYB12, and AtMYB111). AtMYB12 and AtMYB111 can also form part of MYB-bHLH heterodimers that activate transcription of *AtCHS* and *AtF3H* (Deng and Lu 2017; Li and Zachgo 2013). *A. thaliana* LBGs (*AtDFR, AtANS* and *AtUFGT*) can be regulated by several MYB-bHLH-WDR (MBW) complexes, the MYB components of which are subgroup 6 R2R3 MYBs (Stracke et al. 2001). *Antirrhinum majus* has three R2R3 MYBs that

activate floral anthocyanin production through interactions with one of two bHLH proteins. One of the bHLH proteins is necessary for expression of LBGs, in *A. majus* this regulatory grouping includes *F3H, DFR, ANS/LDOX*, and *UFGT*. Different combinations of EBGs and LBGs are regulated by the *A. majus* R2R3-MYB proteins ROSEA 1 (*F3H, FLS, F3´H, DFR, ANS/LDOX, UFGT*), ROSEA 2 (*CHI* and *F3'H*) and VENOSA (*CHI, F3H, FLS, F3´H, ANS/LDOX, UFGT*) (Schwinn et al., 2006). In red grape cultivars there are two subgroup 6 R2R3 MYBs that regulate only the final stages of the anthocyanin pathway – the glycosylation and acylation steps, activating enzymes including UFGT and 3AT (Kobayashi et al. 2002; Matus et al., 2017; Rinaldo et al., 2015; Walker et al., 2007). The affinity of transcription factors for regulating different anthocyanin genes can vary (Petroni and Tonelli 2011). The *N. tabacum* MBW complex (with MYB NtAN2 and bHLH NtAN1), for example, strongly activates LBGs and to a lesser extent EBGs (Bai et al. 2011; Pattanaik et al. 2010). In the monocots *Zea mays* (Petroni and Tonelli 2011), *Oncidium* (Schwinn et al. 2016), and *Lilium spp.* (Lai et al. 2012) there is no clear distinction in regulation, with the same MBW complexes regulating the EBGs and LBGs. This regulatory pattern is also found in the eudicot *Ipomoea nil* (Morita et al. 2006). Subgroup 6 R2R3-MYB transcription factors are strongly associated with regulation of late anthocyanin synthesis genes as part of MYB-bHLH-WDR complexes across many systems (Albert et al. 2011; Dubos et al. 2010; Lai et al. 2012; Martins et al. 2013; Yuan et al. 2014).

Accumulation of anthocyanin pigment within specific petal regions can be regulated by the MBW complexes that activate anthocyanin synthesis genes (e.g. Cooley and Willis, 2009; Thomas et al. 2009; Yamagishi 2014). In pansies (*Viola* x *wittrockiana* Gams.) petal blotches result from upregulation of *F3'5'H*, *DFR*, and *ANS* in spotted tissue (Li et al. 2014), and in *Lilium* species 'splatter' spot formation results from strong upregulation of *DFR* and *ANS* in these regions - with moderate increases in *CHS*, *CHI*, and *F3H* expression (Yamagishi et al. 2014). In *Clarkia gracilis* the activation of a spot specific *DFR* allele, by a MYB expressed within the spot region, leads to cyanidin and peonidin pigmentation of the spot only. Upregulation of *F3'5'H* and a second *DFR* allele at a later developmental stage produces malvidin background pigmentation in the rest of the petal (Martins et al. 2013). Post-transcriptional gene silencing is another mechanism by which pigment is spatially restricted. Silencing of *CHS* causes large reductions in *CHS* transcript levels in the white regions of bicolour anthocyanin pigmented flowers in *Petunia hybrida* (Koseki et al. 2005; Morita et al. 2012; Saito et al. 2006), *Camellia japonica* (Tateishi et al. 2010), and *Dahlia variabilis* (Ohno et al. 2011). Other factors besides alterations involving anthocyanin synthesis enzymes can cause differences in anthocyanin colouration. In *Tulipa gesneriana*, for example, the formation of blue spots on a purple petal background results from anthocyanins complexing with metals within the spot. This complexing results from spot-specific transcription of a vacuolar iron transporter and suppression of a gene causing Fe storage (Momonoi et al. 2009; Shoji et al. 2007; Shoji et al. 2010). White regions of *Mimulus lewisii* petals preferentially express a gene encoding the enzyme FLS that uses dihydroflavonols to form flavonols, causing diversion of the dihydroflavonol substrate away from anthocyanin production and into the flavonol pathway (Yuan et al. 2016). Evidently, there are several mechanisms by which differential colouration is produced in specific petal regions, resulting in petal anthocyanin patterning.

Each gene encoding a type of anthocyanin synthesis enzyme can be present as one copy or multiple copies depending on the plant species (Streisfeld and Rausher 2009). When a single copy is present, or one copy is expressed in most tissues and the additional copies have more restricted expression patterns, loss of gene expression or functioning may lead to deleterious pleiotropy (Durbin et al. 2003; Inagaki et al. 1999; Koes et al. 1989; Streisfeld and Rausher 2009). In *Ipomoea purpurea* a purple floral

phenotype is predominant but white flowered variants do occur within the same populations, with the latter caused by a transposon insertion into a *CHS* locus, preventing transcription of a functional gene copy (Habu et al. 1998; Johzuka-Hisatomi et al. 1999). This reduction in CHS activity in floral tissue likely reduces production of other flavonoids besides anthocyanins. Negative pleiotropic effects of the white floral variant have been experimentally determined and are thought to contribute toward the rarity of the allele in populations (Coberly and Rausher 2003, 2008; Fehr and Rausher 2004; Habu et al. 1998). If multiple types of pigment are produced from more than one branch of the anthocyanin pathway, loss of function of one enzyme can cause a change in anthocyanin composition rather than a loss of anthocyanin phenotype. In *Iochroma*, deletion of an F3'5'H encoding gene and downregulation of F3'H expression caused a shift to pelargonidin pigmentation producing red coloured flowers in *Iochroma gesnerioides*, as opposed to the blue flowers of many *Iochroma* species (Smith and Rausher 2011). Evolutionary comparisons exploring how developmental processes are modified to produce differences in pigmentation also contribute to mechanistic understanding.

Ideally, genes of interest are manipulated in the host system to gain insight into functionality. This would enable examination of the mechanisms of anthocyanin regulation and identification of downstream targets in the anthocyanin synthesis pathway. Upregulating and repressing individual genes within the regulatory network would provide information on whether each candidate is sufficient and/or necessary for anthocyanin production, while resultant phenotypes provide further information on gene functioning. In *G. diffusa* a transformation protocol is still being developed, as such the transformation work used to determine the role of *GdMYB8* transcription factors in anthocyanin regulation was undertaken in the heterologous system *N. tabacum*. This species has been used in many studies across several systems for assessing the function of candidate anthocyanin regulatory genes (Elomaa et al. 2003; Mooney et al. 1995; Nakatsuka et al. 2013; Pandey et al. 2014).

In the previous chapter, three subgroup 6 R2R3 MYB transcription factors were identified as potential regulators of petal spot anthocyanin pigmentation in *G. diffusa*. This chapter aims to determine whether specific anthocyanin synthesis enzymes are good candidates for regulation by these GdMYB8 proteins, and whether GdMYB8 proteins are sufficient to regulate anthocyanin synthesis in a heterologous system. The anthocyanin synthesis genes have been characterised within other Asteraceae including *Chrysanthemum* (He et al. 2013) and *Gerbera hybrida* (e.g. Helariutta et al. 1995; Helariutta et al. 1993). In *Chrysanthemum* two MYBs are likely regulators of *DFR,* and in *G. hybrida* GMYB10 strongly induces transcription of the LBGs (Laitinen et al. 2008; Liu et al. 2015a). The focal genes chosen here were *anthocyanidin synthase* (*ANS*), *dihydroflavonol 4-reductase* (*DFR*), and *malonyl transferase* (*MAT*). *ANS* and *DFR* were chosen because they encode late biosynthesis enzymes and so form products committed to anthocyanin (and for *DFR* proanthocyanidin) production. *MAT* encodes an acyl transferase and acts downstream of *ANS* and *DFR* catalysing a reaction that adds malonyl groups to anthocyanins. It was selected as a potentially interesting candidate because a large proportion of anthocyanins in *G. diffusa* spotted petal regions contain malonyl residues, while anthocyanins from plain petal tissue do not (or at very low quantities). As such, *GdMAT* may function in a spot-specific manner. These three enzymes were first identified and characterised within *G. diffusa* through gene hunting and RNA-seq data (Walker 2012, Kellenberger unpublished). Expression analyses were used to determine whether the genes encoding these enzymes were upregulated within developing petal spots. Upregulation would suggest that these enzymes may have a role in spot development, along with correlative evidence for regulation by GdMYB8 proteins. To establish

whether GdMYB8 proteins were sufficient to regulate anthocyanin biosynthesis enzymes, *GdMYB8* genes were ectopically expressed in the heterologous host *Nicotiana tabacum*.

Enzyme	Abbreviation
Chalcone synthase	CHS
Chalcone isomerase	CHI
Flavonone 3-hydroxylase	F3H
Flavanone 3’ hydroxylase	F3’H
Flavanone 3’, 5’ hydroxylase	F3’5’H
Dihydroflavonol 4-reductase	DFR
Anthocyanidin synthase	ANS
UDP-3-O-glucosyltransferase	UFGT
Malonyl transferase	MAT
Flavonol synthase	FLS
Anthocyanin 3-O-glucoside-6”-O-acyltransferase	3AT

Table 5.1. Flavonoid/ anthocyanin synthesis pathway enzymes abbreviated in the main text.

5.2 Methods

5.2.1 Characterising genes encoding *G. diffusa* anthocyanin synthesis enzymes

Genes encoding anthocyanidin synthase (ANS), dihydroflavonol 4-reductase (DFR), and malonyl transferase (MAT) were investigated as potential downstream targets of the GdMYB8 proteins. *GdANS* and *GdDFR* were characterised in Cal and Spring only, whereas *GdMAT* was also characterised in Stein.

From a recent RNA-seq analysis (Kellenberger unpublished), 5 malonyl transferase genes/ alleles were identified. Of these, 4 contained the conserved motif (YFGNC(A)) of the subfamily for anthocyanin malonyl transferases (Unno et al. 2007) and 1 of these (here referred to as *GdMAT1*) was upregulated in spotted petal tissues compared to plain tissue. As such, this variant was selected as a good candidate for anthocyanin production in the spot. The full length of this *GdMAT1* coding sequence was present within the RNA-seq data and so primer pairs were designed (Section 2.3.1) from the UTRs to amplify the full *G. diffusa MAT1* coding sequence.

Partial sequences of *GdANS* and *GdDFR* genes had previously been obtained from a 454 *G. diffusa* transcriptome of petal tissue (Walker 2012). There was a single copy of *ANS* within this transcriptome and multiple *DFR* variants. Following characterisation of these genes through gene hunting, a more recent RNA-seq analysis was conducted (Kellenberger unpublished) that verified the findings. 3'RACE (Section 2.3.5) was used to isolate the 3'UTR sequences of both types of genes, with forward primers designed using the partial gene sequences from the transcriptome. Genome walking was used to find the 5' regions of the genes. For *ANS*, genome walking (Section 2.3.3) successfully revealed the start of the coding sequence, including the start codon and a section of 5'UTR. The 5' region of *DFR* was difficult to obtain through genome walking due to a large repetitive section of DNA within an intron, so degenerate primers were designed (Section 2.3.4) to circumvent this issue and used to successfully isolate a large section at the beginning of the gene. Genome walking was then used, with primers designed from this newly isolated region, to obtain the full length of the gene. All PCRs were conducted with the proof-reading enzyme Phusion (NEB) (Section 2.3.2). For *GdANS*, *GdDFR*, and *GdMAT,* primer pairs (Appendix 2) were then designed in the 5'UTR and 3'UTR and the full length of each gene was amplified in one PCR from the gDNA and cDNA of multiple *G. diffusa* individuals, cloned, and sent for Sanger sequencing (Section 2.4). The characterisation was conducted in Spring and then the same primers were used to amplify genes in Cal and Stein.

5.2.2 Floral expression patterns of *GdANS*, *GdDFR*, and *GdMAT1*

qRT-PCR was used to determine the expression levels of *GdANS*, *GdDFR*, and *GdMAT1* in spotted and plain ray floret tissue at two developmental stages during spot development. The general methods for the tissue preparation, RNA extraction, and cDNA synthesis are in Section 2.2 and for qRT-PCR in Section 2.7. The details of collection of the specific *G. diffusa* ray floret tissue used in this analysis are outlined in Section 4.2.6. Primers for qRT-PCR (Appendix 2) were designed based on the single copy of *GdANS* identified, to be specific to *GdMAT1*, and to bind to all variants of *GdDFR* – as the *GdDFR* variants were too similar in sequence to design variant-specific primers. As an exhaustive search for all copies of these genes in *G. diffusa* was not conducted, while primers were designed to be specific to the particular enzyme-coding sequences (through comparisons with other Asteraceae sequences), it is unknown how many genes were amplified in the qRT-PCR for each type of enzyme.

5.2.3 Stable transformations of *GdMYB8* genes into *Nicotiana tabacum*

To provide insight into the functionality, the 3 GdMYB8 proteins (GdMYB8a, GdMYB8b, and GdMYB8c) encoded by genes found to be upregulated in spotted petal tissue (Section 4.3.6) were stably expressed individually in *Nicotiana tabacum*. The coding sequence of each *GdMYB8* candidate was cloned from the *G. diffusa* Spring morphotype into a pGreen construct and placed under the control of a constitutive double 35S CaMV promoter and a 35S terminator sequence. This construct was used to stably transform *N. tabacum*, with several independent insertion lines produced for each gene (*GdMYB8a* n = 4, *GdMYB8b* n = 8, *GdMYB8c* n = 8).

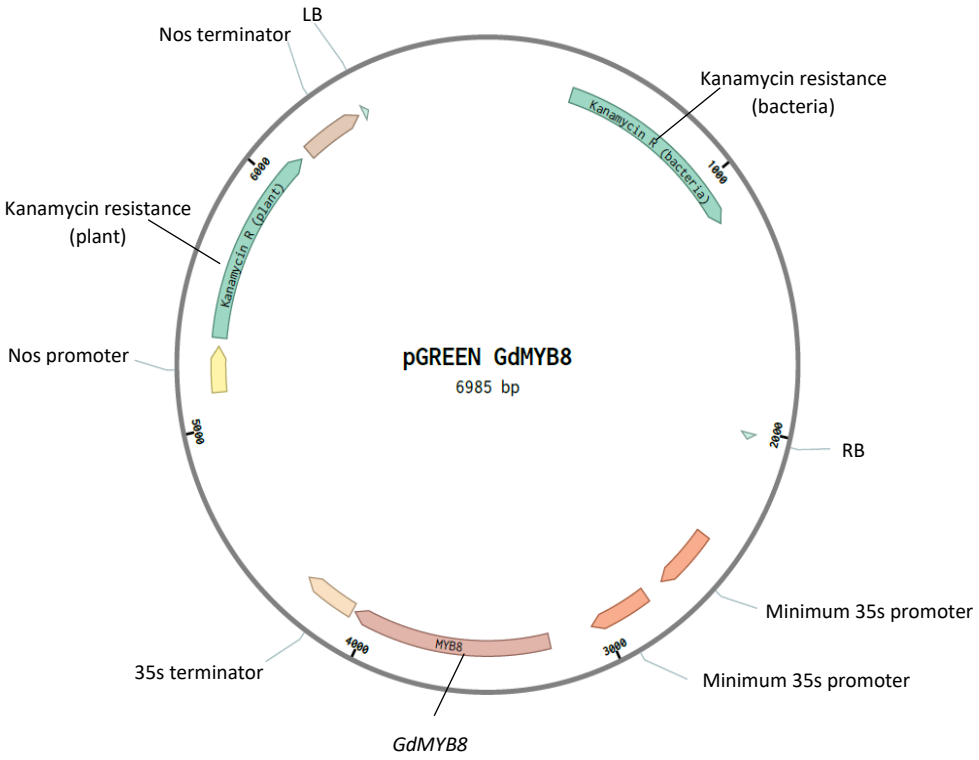


Figure 5.1. Diagram of the pGreen vector used in stable *Nicotiana tabacum* transformations, inducing constitutive expression of GdMYB8 proteins. Three identical vectors were constructed, differing only in the gene they contained: either *GdMYB8a*, *GdMYB8b*, or *GdMYB8c*.

Producing stable transformants of *N. tabacum*

The transformation process for *GdMYB8a* was completed previously by Mellers (2016) and seeds from the T_0 of this transformation were grown in parallel with seeds produced by the T_0 generation from the current transformation experiment. Vector construction is detailed in Section 2.5.2 and the completed vector is illustrated in Figure 5.1. Stable transformation of these vectors into *N. tabacum* and subsequent tissue culture is detailed in Section 2.6. Once potted in soil, leaf tissue was removed from each plant, flash frozen, and used for genotyping – along with wild type controls. RNA was extracted, cDNA synthesised, and PCR conducted to determine whether the transgene was being expressed (Section 2.2). Eight transformed *N. tabacum* plants (8 independent lines) for each of *GdMYB8b* and *GdMYB8c* were retained and allowed to self-fertilise and set seed. These seeds were grown along with the 4 lines of *GdMYB8a N. tabacum* available and 3 wild type plants. This T_1 generation was genotyped to check for transgene expression and phenotyped to determine GdMYB8

function within this heterologous system. A single T_1 generation plant from each line had leaves and flowers removed, and these were photographed, along with wild type equivalents, using a Nikon coolpix P520 camera. Quantification of anthocyanin content within the flowers of transgenic *N. tabacum* was conducted on mature flower anthers, sepals, and petals. The stage at which these samples were taken, was standardised based on the developmental series in Dek et al. (2017) (Fig 5.2). Three to five flowers from each tobacco plant were dissected and the different tissue types from each flower were individually flash frozen in liquid nitrogen, weighed, and anthocyanin was extracted and quantified as detailed in Section 4.2.2. Unfortunately, this could not be conducted across all lines for *GdMYb8b* and *GdMYB8c* so n = 5 and n = 6 lines were used, respectively.

Figure 5.2. The developmental stages of wild type *N. tabacum* flowers (modified from Dek et al. (2017)). The red boxes indicate the stages used in the analyses: (a) for qRT-PCR samples (b) for anthocyanin extraction samples.

Determining *GdMYB8* expression levels in *N. tabacum*

qRT-PCR was used to determine the expression levels of *GdMYB8* transgenes and potential downstream targets - the genes encoding the anthocyanin synthesis enzymes *NtANS*, *NtDFR*, and *NtMAT*. The floral expression patterns of the *G. diffusa* homologues of these genes were also investigated within this chapter. qRT-PCR procedures and primer efficiency tests are detailed in Section 2.7. Petal tissue was harvested at between 14:30 - 15:45 at developmental stage one (as defined in Dek et al. (2017) (Fig 5.2), petals from multiple flowers within an individual were pooled. *N. tabacum* qRT-PCR reference genes were selected by examining the literature. Schmidt and Delaney (2010) investigated the stability of several genes across a range of tissue, including floral tissue. They recommended use of the 2 most stable genes from these analyses, demonstrating that 2 reference genes are required for accurate normalisation. Since then the *N. tabacum* genome has become available, so these recommended reference genes were blasted against the *N. tabacum* genome to check that only 1 gene was in fact being amplified. This led to exclusion of one of the recommended genes, which was replaced by the 3rd most stable gene from Schmidt and Delaney (2010) following these checks. As such, *N. tabacum Elongation Factor 1* (*NtEF-1*) and *Ubiquitin C* (*NtUBC*) were used as reference genes, these were also used as reference genes in *N. tabacum* by Divya et al. (2019).

NtDFR and *NtANS* primer sequence were obtained from Yamagishi et al. (2014) and modified so that the annealing temperature was optimal for qRT-PCR using the Luna enzyme. Modified primers were blasted against the *N. tabacum* genome to ensure only *DFR* or *ANS* genes would be amplified. Transgenic and wild type flowers were sent for LC-MS analysis to investigate how anthocyanin composition in transgenics deviated from wild type, with a particular interest in determining whether anthocyanins with malonyl residues were present in transgenics. Unfortunately, due to COVID disruption, these samples could not be analysed. As an alternative, malonyl transferase expression

was investigated through qRT-PCR. *NtMAT* primers were designed based on those in Taguchi et al. (2005), and blasted against the *N. tabacum* genome to determine whether they were specific to malonyl transferases. Based on these assessments, and primer efficiency tests, alterations were made to produce the *NtMAT* primers used here (Appendix 2). *GdMYB8* primers from *G. diffusa* qRT-PCR could not be used here because they included a section of the 3'UTR, but only the coding sequence was transformed into tobacco. Primers were designed to be specific to *G. diffusa* subgroup 6 MYBs, while minimising amplification of *N. tabacum* subgroup 6 MYB genes. It was not possible to design a suitable primer to amplify all 3 *GdMYB8* genes, so 2 primers were designed: one to amplify *GdMYB8a* and the second to amplify *GdMYB8b* and *GdMYB8c* (Appendix 2). Programs used for statistical tests and graphical representations of the data are detailed in Section 2.7.4. A subset of lines was used in qRT-PCR, chosen to represent the full diversity of anthocyanin phenotype within transformants of each transgene. In total complete sets of pigment quantification and qRT-PCR data were obtained for 4 lines of *GdMYB8a* and *GdMYB8c* transformants and 5 lines of *GdMYB8b* transformants. qRT-PCR was conducted in an additional line of *GdMYB8c* and 2 lines of *GdMYB8b*.

Figure 5.3. Gel electrophoresis images demonstrating that the relevant *GdMYB8* gene is being expressed in each transgenic tobacco line in the T_1 generation. WT is wild type *N. tabacum* and the negative control for the PCR is indicated (-). Primers used to amplify *GdMYB8b* and *GdMYB8c* transgenes are the same, hence there is a single negative control and wild type sample present. Expected band sizes were as follows: *GdMYB8a* 478bp, *GdMYB8b* 520bp, and *GdMYB8c* 549bp.

5.3 Results

5.3.1 Characterising the genes encoding *G. diffusa* anthocyanin synthesis enzymes

Anthocyanidin synthase (*ANS*), *dihydroflavonol 4-reductase* (*DFR*), and *malonyl transferase* (*MAT*) encode enzymes within the anthocyanin synthesis pathway. These genes were characterised in *G. diffusa* as a first step toward determining whether they are downstream targets of GdMYB8 proteins. The characterisation was not exhaustive, but results were corroborated by ensuring that gene hunting identified all of the variants present in the transcriptomes (Walker 2012, Kellenberger unpublished). The structure of each enzyme encoding gene is illustrated in Figure 5.4.

A single copy of *GdANS* was found, with full length gDNA of 2484bp containing a single large (1266bp) intron. The gDNA and cDNA of *GdANS* was characterised in 4 individuals, and minimal nucleotide differences were found between alleles within and across individuals. The *GdANS* coding sequences differed at 33 nucleotide positions, all of which were synonymous. As such, a single *GdANS* protein sequence was obtained that was 405 amino acids in length.

The characterisation of *malonyl transferase* focused on a single variant (*GdMAT1*) that was upregulated in spotted compared to non-spotted petal regions within a *G. diffusa* transcriptome (Kellenberger unpublished). This gene contained no intron and GdMAT1 is 463 amino acids in length. *GdMAT1* is a gene, rather than an allelic variant, as two copies were found within single plants. Across the 6 individuals in which *GdMAT1* was characterised there were 16 nucleotide differences resulting in non-synonymous changes to the protein sequence, these amino acid differences are listed in Appendix 5.

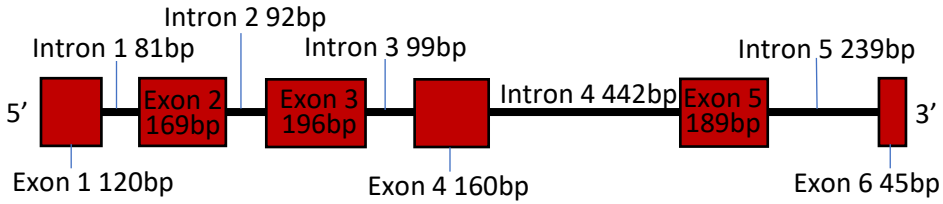
Multiple *GdDFR* genes were found, but it is unclear how many copies are present within the genome and expressed within petal tissue of *G. diffusa*. Full length gDNA was between 1775 - 1832bp and the coding sequence, amplified from cDNA, was 879bp. *GdDFR* consisted of 6 exons and 5 introns (Fig 5.4). Between and within individuals there was high sequence variability within intron 4 and intron 5. In 2 individuals, 5 different *GdDFR* sequences were found indicating that there are at least 3 *DFR* genes present within *G. diffusa*. Across all *GdDFR* sequences amplified from multiple individuals, the coding regions were relatively highly conserved (28 SNPs, 17 amino acids differences - detailed in Appendix 5). Within the *DFR* region thought to confer substrate specificity (Johnson et al. 2001), there were 3 positions where amino acids differed between *GdDFR* sequences. In one individual, 2 of the gDNA *GdDFR* sequences were very similar (13 SNPs, and a single amino acid difference), but one variant had a stop codon in exon 1 and a second stop codon in exon 2.

Due to the similarity in coding sequences, it was difficult to establish if all *GdDFR* variants were present in petal cDNA. 3'UTRs were very similar and could not be successfully obtained for all of the specific sequences found within each individual. Exploration for 3'UTR sequences (when using gDNA) led to the discovery of short sections of *GdDFR* that diverged from all previously characterised sequences. The 'variants' obtained were tentatively categorised into A-D based on minimal nucleotide differences in coding regions, and all of these variants were found within cDNA. *GdDFR* was amplified in cDNA using primers located in the 5'UTR and 3'UTR. The products obtained from PCR and cloning were of several different lengths. Within a single individual, for example, products of 581bp (with a stop codon at 545bp), 737bp (with a stop codon at 350bp), and 270bp (with a stop codon at 235bp) were obtained – none of these 3 cDNA amplicons had identical nucleotide sequences. Additional lengths of *GdDFR* coding sequences were obtained from cDNA found in other individuals. All of these cDNA sequences

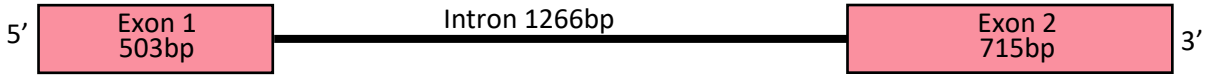
had conserved amino acids, aligning with the 5' end of the gene, but with premature stop codons. One cDNA sequence amplified had a stop codon at 305bp but was 1925bp in length. The coding sequence was comprised of exon 1 and exon 2 but, following the stop codon, the remaining cDNA sequence contained all of the additional 4 introns found within the gDNA.

For the purposes of this study, full characterisation of all copies of these enzymes was not required. Here the presence of the catalytic triad of amino acids that perform covalent catalysis (Dodson and Wlodawer 1998) was used as a proxy to determine whether truncated GdDFR proteins were likely to be functional. The catalytic triad (amino acid and position in protein sequence: S-129, Y-164, K-167) (Petit et al. 2007) was present in all cDNAs where the coding sequence was longer than 501bp. A catalytic triad of amino acids has also been identified for *ANS* and was present within *GdANS* (E-142, K-215, N-217) (Wilmouth et al. 2002).

Dihydroflavonol 4-reductase (GdDFR)



Anthocyanidin synthase (GdANS)



Malonyl transferase (GdMAT1)

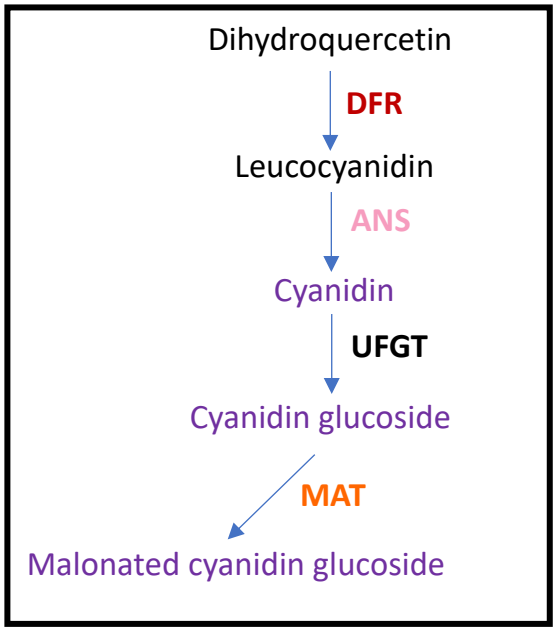


Figure 5.4. Schematic of the genomic DNA of *G. diffusa* genes encoding the anthocyanin synthesis enzymes dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), and malonyl transferase (MAT1). Black lines represent introns, and exons are represented by coloured blocks. The final stages of the anthocyanin pathway involving the late biosynthetic genes, and the products they synthesis, are illustrated in the black box. Enzymes are depicted in bold next to the arrows and compound names are in black if they are colourless and purple if they are coloured. UFGT is UDP-3-O-glucosyltransferase.

5.3.2 *GdANS*, *GdDFR*, and *GdMAT1* are upregulated in developing petal spots

Note. The following abbreviations are used: Cal spotted (Sp) and plain (Tp) ray floret petal segments. Spring whole spotted ray floret petals (Sr) and whole plain (Pr) ray floret petals (Pr). Developmental stage 1 (1) and 2 (2) Diagrams of petal sections are in Fig 5.5.

To determine whether the anthocyanin synthesis enzymes were upregulated in developing petal spots, qRT-PCR was conducted on spotted and plain petal tissue at two developmental stages in Cal and Spring (Fig 5.5). *GdMAT1* expression was also investigated in Stein. Due to the variability and ambiguity in *GdDFR* genes, qRT-PCR *DFR* primers were designed to amplify all cDNA that contained the catalytic triad of amino acids. In both Cal and Spring, *GdANS* and *GdDFR* were significantly more highly expressed in spotted compared to plain petal tissue at developmental stage 1 (Cal Sp1 – Tp1: *GdANS* t = 4.62 p = 0.002, *GdDFR* t = 7.64 p = 0.0001; Spring Sr1 – Pr1: *GdANS* t = -7.073 p = 0.0002, *GdDFR* t = -13.25 p <0.0001). In plain petal tissue of Spring and Cal there was a significant increase in *GdANS* and *GdDFR* expression at developmental stage 2 (Cal Tp1 – Tp2: *GdANS* t = -10.67 p <0.0001, *GdDFR* t = -23.6 p <0.0001; Spring Pr1 – Pr2: *GdANS* t =-10.86 p <0.0001, *GdDFR* t =-17.15 p <0.0001). Due to this increase in expression, there was no difference in expression levels of *GdANS* between spotted and plain petal tissue at developmental stage 2, but *GdDFR* was upregulated in spotted petal tissue in Cal (Sp2 – Tp2 *GdDFR* t = 5.56, p = 0.002) and Spring (Sr2 – Pr2 t = -2.79, p = 0.024).

Expression levels of *GdANS* and *GdDFR* were compared between equivalent developmental stages and tissue within a morphotype (Fig 5.5). In Cal *GdANS* had consistently higher expression than *GdDFR* (Sp1 t = 4.01 p = 0.011, Tp1 t = 4.45 p = 0.001, Sp2 t = 3.48 p = 0.004, Tp2 t = 4.00 p = 0.001) and the same pattern was seen in Spring (Pr1 t = 3.52 p = 0.004, Sr2 t = 3.11 p = 0.009, Pr2 t = 2.97 p = 0.011), with the exception of spotted tissue at developmental stage 1.

In Cal and Spring *GdMAT1* was upregulated at both developmental stages in spotted petal tissue compared to plain tissue (Fig 5.5) (Cal Sp1 – Tp1 t = 4.27 p = 0.008, Sp2 – Tp2 t = 2.76 p = 0.048; Spring Sp1 – Tp1 t = -5.24 p = 0.002, Sp2 – Tp2 t = -4.31 p = 0.005). *GdMAT1* was expressed at much lower levels overall than *GdANS* and *GdDFR* (Fig 5.5, Fig 5.6). In Stein, which has no petal spots, *GdMAT1* expression was also very low and apparently 0 in some biological replicates (mean ± s.e: D1 0.01 ± 0.004, D2 0.012 ± 0.01).

Fig 5.6 demonstrates the relative expression levels of the *G. diffusa* anthocyanin biosynthesis enzymes investigated alongside the *GdMYB8* transcription factor expression levels (detailed in Section 4.3.6).

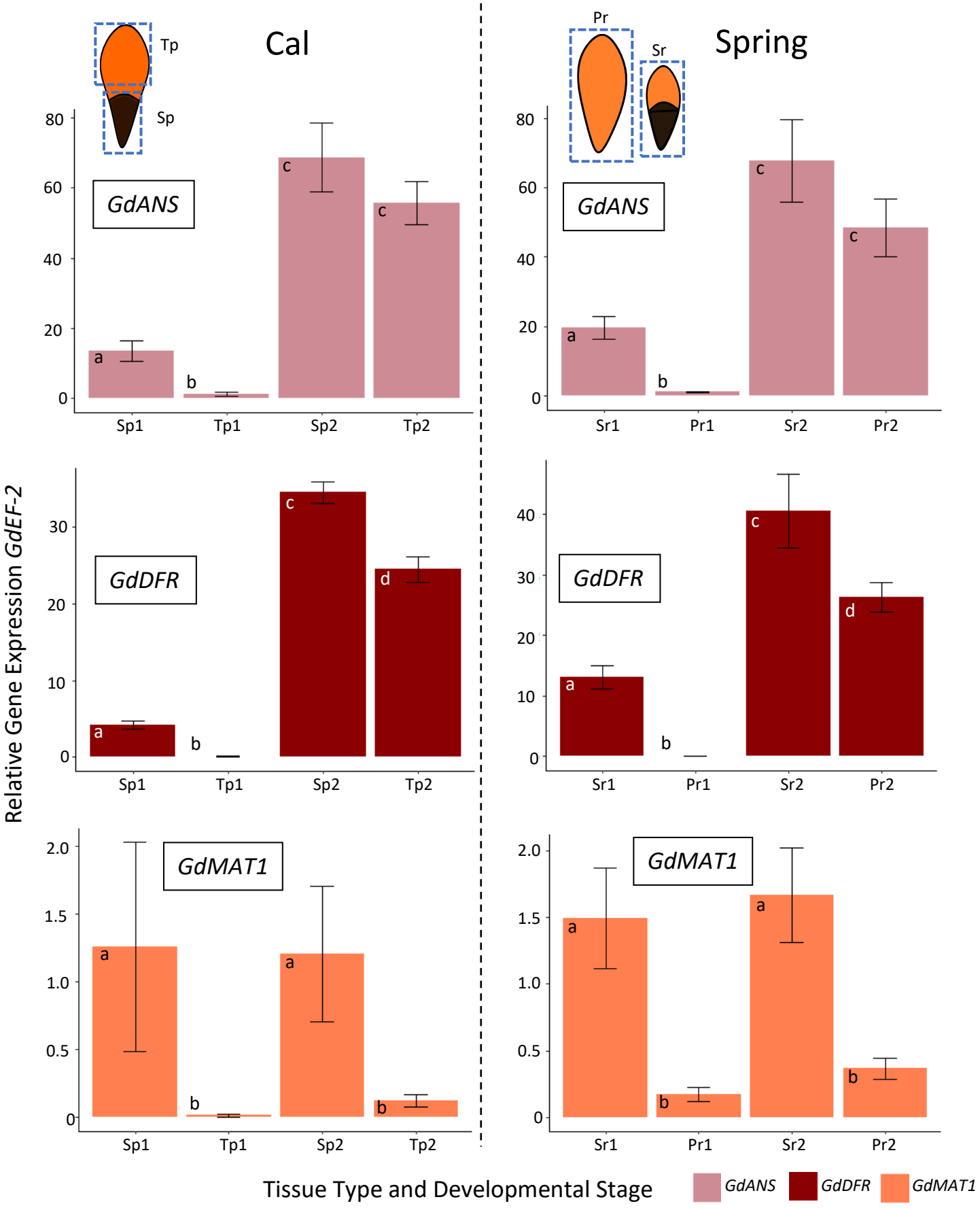


Figure 5.5. qRT-PCR results showing the relative expression levels of genes encoding the anthocyanin synthesis enzymes anthocyanidin synthase (ANS), dihydroflavonol 4-reductase (DFR) and malonyl transferase (MAT1) in *G. diffusa* morphotypes Spring and Cal. Two developmental stages were used in Cal and Spring morphotypes: Cal spotted (Sp) and plain (Tp) ray floret petal tissue (left-hand column) and Spring whole spotted ray floret petals (Sr) and whole plain ray floret petals (Pr) (right-hand column). The tissue segments used are indicated in the ray floret diagrams for each morphotype. Error bars represent the mean ± s.e. Within a graph, tissues with statistically significant differences ($p \leq 0.05$) in expression levels do not share a letter. The sample size for each tissue type was n = 3.

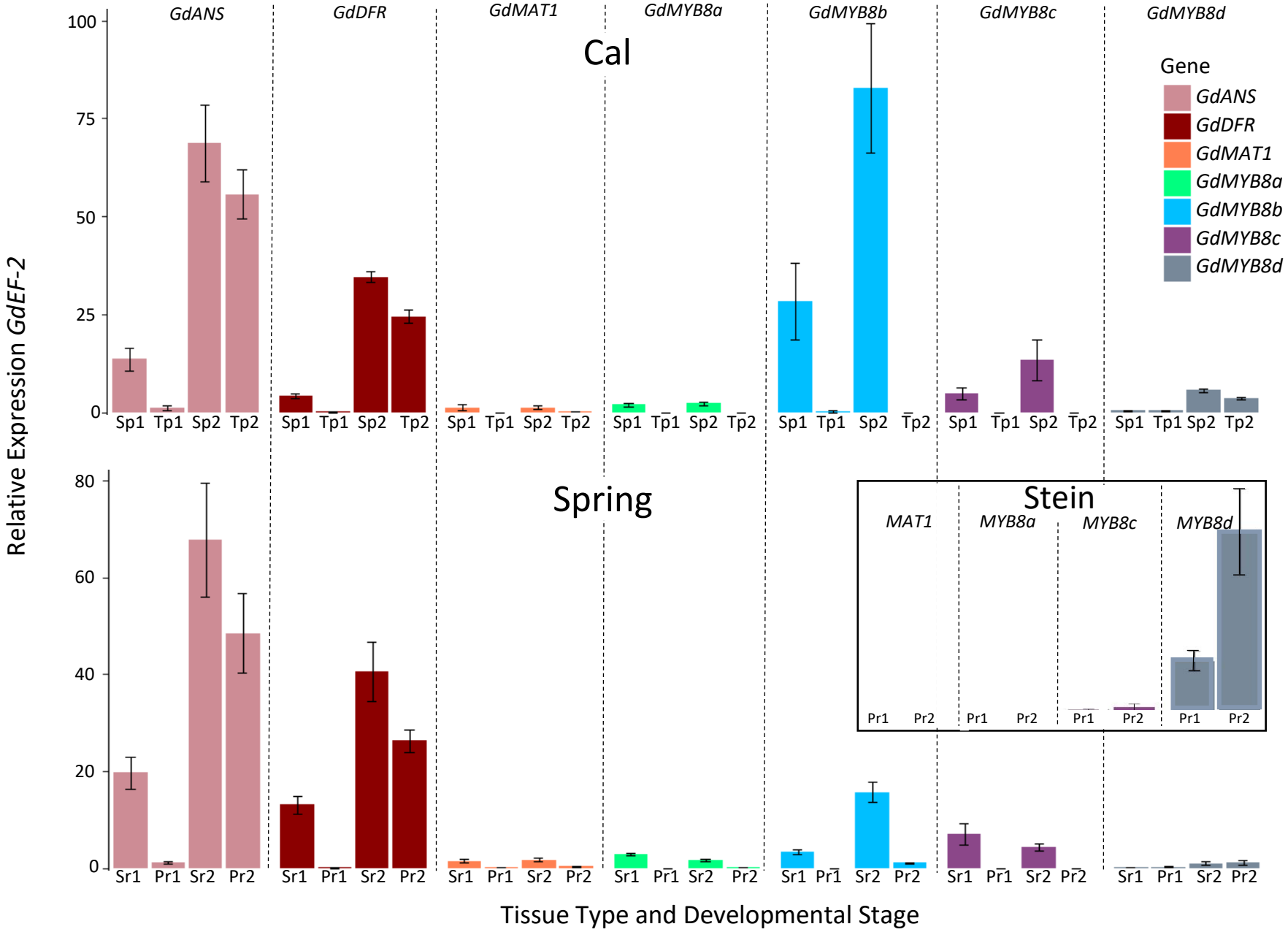


Figure 5.6. The relative expression of all *GdMYB8* genes and focal anthocyanin synthesis enzymes at two developmental stages across Cal (top row), Spring (bottom row), and Stein (in the black box), in spotted tissue (Sp/ Sr) and plain tissue (Tp/ Pr). Error bars represent the mean ± s.e.

5.3.3 GdMYB8 proteins activate anthocyanin synthesis in a heterologous system

Each *GdMYB8* that was upregulated in *G. diffusa* spotted petal tissue (*GdMYB8a*, *GdMYB8b*, *GdMYB8c*) was stably transformed into *N. tabacum* on a constitutive promoter. These transformed plants were used to assess whether GdMYB8 proteins were capable of regulating anthocyanin synthesis. Transformed plants (T_0) were grown to maturity, allowed to self-fertilise, and the resulting seeds were harvested and grown (T_1). Genotypic and phenotypic data were collected from the T_1 generation. Stable transgene expression was verified through PCR, that amplified transgenes from floral bud cDNA, and wild type floral bud cDNA was tested in parallel and showed no amplification (Fig 5.3). Across all lines transformed with 1 of the 3 *GdMYB8* genes, constitutive expression of the transgene was sufficient to induce anthocyanin production.

Fig 5.7i. Flowers and leaves from *N. tabacum* plants transformed with *GdMYB8a* on a constitutive promoter. Within each photograph, each flower/ leaf represents an independent line. WT indicates a wild type plant for comparative purposes.

Fig 5.7ii. Flowers and leaves from *N. tabacum* plants transformed with *GdMYB8b* on a constitutive promoter. Within each photograph, each flower/ leaf represents an independent line. WT indicates a wild type plant for comparative purposes. Plants from lines with the strongest anthocyanin phenotype often took longer to flower, hence why four lines are represented in flower comparisons and 7 in leaf comparisons.

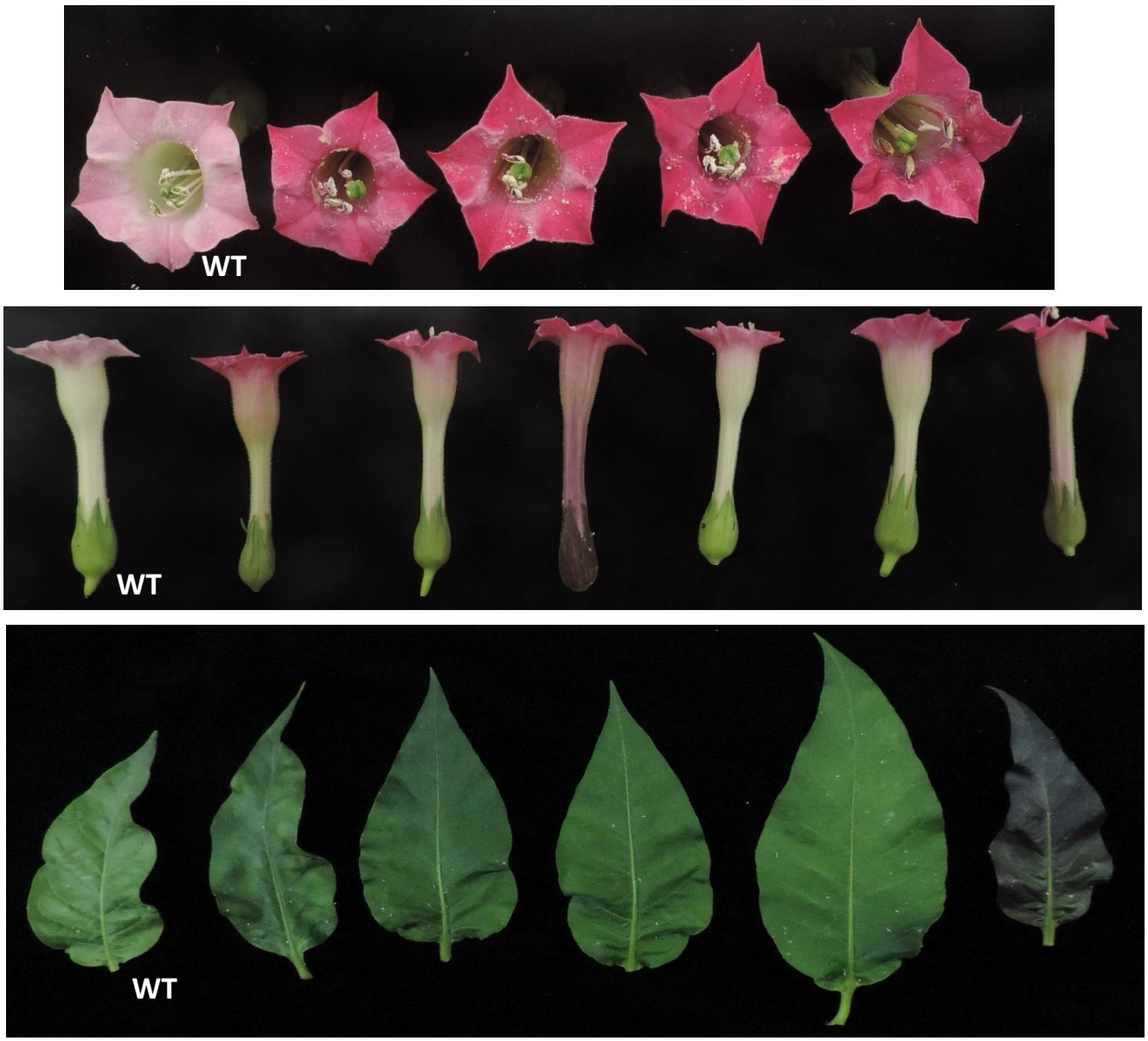


Fig 5.7iii. Flowers and leaves from *N. tabacum* plants transformed with *GdMYB8c* on a constitutive promoter. Within each photograph each flower/ leaf represents an independent line. WT indicates a wild type plant for comparative purposes.

Transformants had stronger anthocyanin phenotypes than wild type plants in petal tissue across all lines. In at least one line from each *GdMYB8* transformant, ectopic anthocyanin production occurred in other tissues (sepals, anthers, and leaves) (Fig 5.7i - iii). In *GdMYB8a* lines, leaf anthocyanin pigmentation was only visible once the leaves began to brown. *GdMYB8b* lines tended to have the strongest anthocyanin phenotypes. The anthocyanin content of anthers, sepals, and petals was measured within transgenic lines, so that variability between lines transformed by the same *GdMYB8* gene could be quantified. A minimum of 3 flowers from each plant (representing one line) was dissected and the mean anthocyanin content of each tissue was taken (Fig 5.8). Across all constructs petal anthocyanin content was significantly higher in *GdMYB8* transformants than wild type petals (*GdMYB8a* $t = 2.82$ $p = 0.03$, *GdMYB8b* $t = 4.58$ $p = 0.002$, *GdMYB8c* $t = 3.74$ $p = 0.006$), but there was no significant difference in anthocyanin content between *GdMYB8* constructs. Anther and sepal anthocyanin content were significantly higher than in wild type plants for *GdMYB8b* (anther: $t = 5.83$ $p = 0.0003$, sepal: $t = 6.91$ $p < 0.0001$) and *GdMYB8c* (anther: $t = 3.58$ $p = 0.01$, sepal: $t = 2.73$ $p = 0.023$) transformants. *GdMYB8a* anther anthocyanin content was also higher than wild type with marginal significance ($t = 2.40$ $p = 0.48$). The sepal anthocyanin content of *GdMYB8b* transformants was significantly greater than that of *GdMYB8a* ($t = -5.72$, $p = 0.0001$) and *GdMYB8c* ($t = 4.74$, $p = 0.001$) transformants. *GdMYB8b* transformants contained significantly more anthocyanin in anthers than *GdMYB8a* ($t = -3.29$, $p = 0.012$) transformants. Relative transgene expression in petal tissue was determined using qRT-PCR in a subset of transformed *N. tabacum* lines (Fig 5.9) (*GdMYB8a* plants $n_{lines} = 4$, *GdMYB8b* plants $n_{lines} = 7$, *GdMYB8c* plants $n_{lines} = 5$). As expected, wild type controls showed no amplification from transgene primers. Transgene expression was generally higher across *GdMYB8b* lines compared to transgene expression in *GdMYB8a* lines and *GdMYB8c* lines. As such, the stronger phenotype seen within *GdMYB8b* transgenic lines could be a consequence of higher transgene expression. The relative role of this factor compared to any innate differences between GdMYB8 proteins cannot be disentangled from these data.

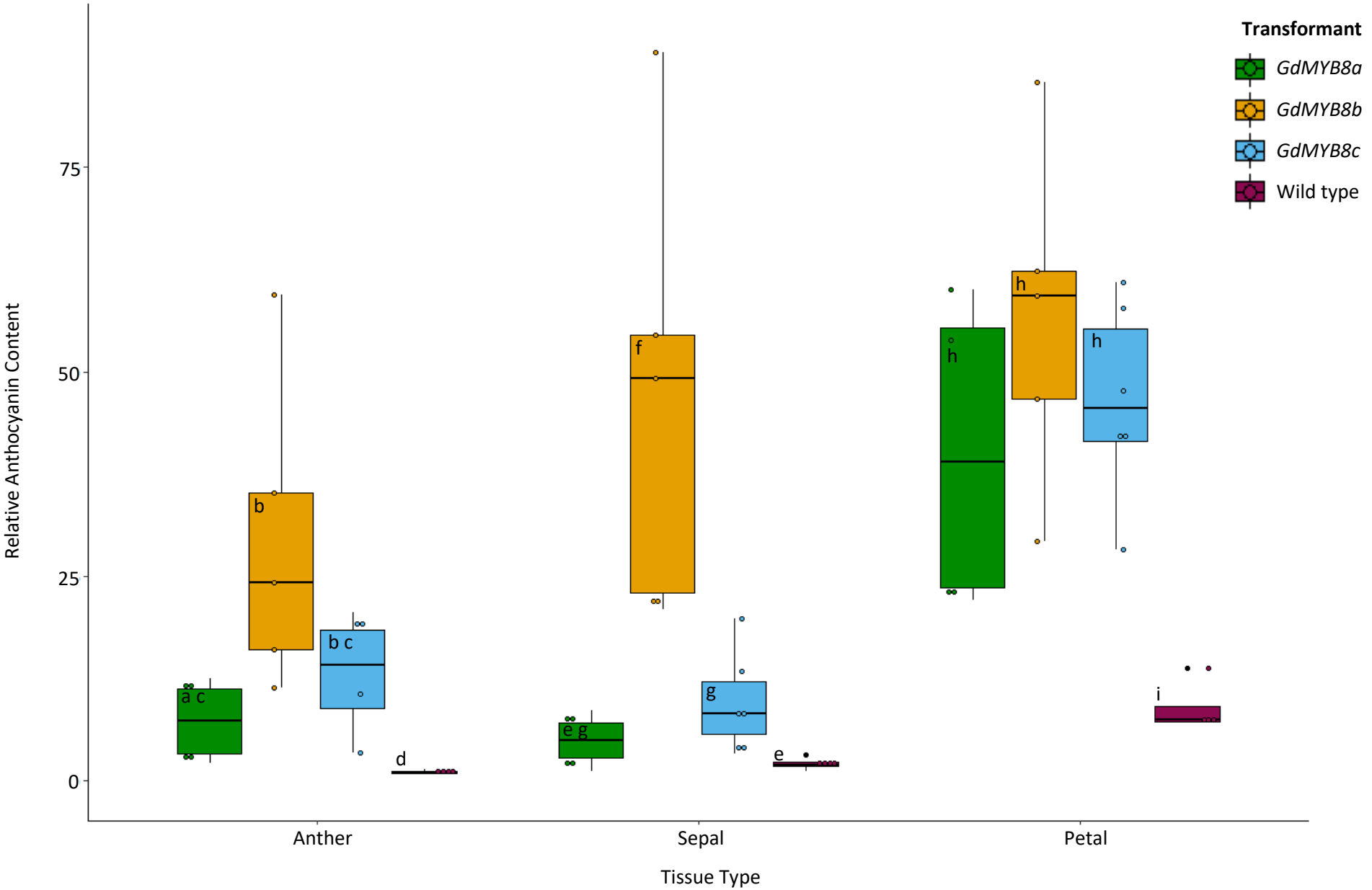


Figure 5.8. The relative anthocyanin concentration within anthers, petals, and sepals of *N. tabacum* T_1 plants. This demonstrates the range of anthocyanin concentrations found between different lines of each set of transformants (*GdMYB8a*, *GdMYB8b*, *GdMYB8c*). The black line in each box indicates the median value and the whiskers are the 25/75% quantile +/- 1.5 *IQR, respectively. Each data point is represented by a black circle and is the mean anthocyanin content in one independent line calculated from 3 - 5 samples (n = 2 for one *GdMYB8c* petal sample). Within a tissue type, transformants with statistically significant differences ($p \leq 0.05$) in relative anthocyanin content do not share a letter.

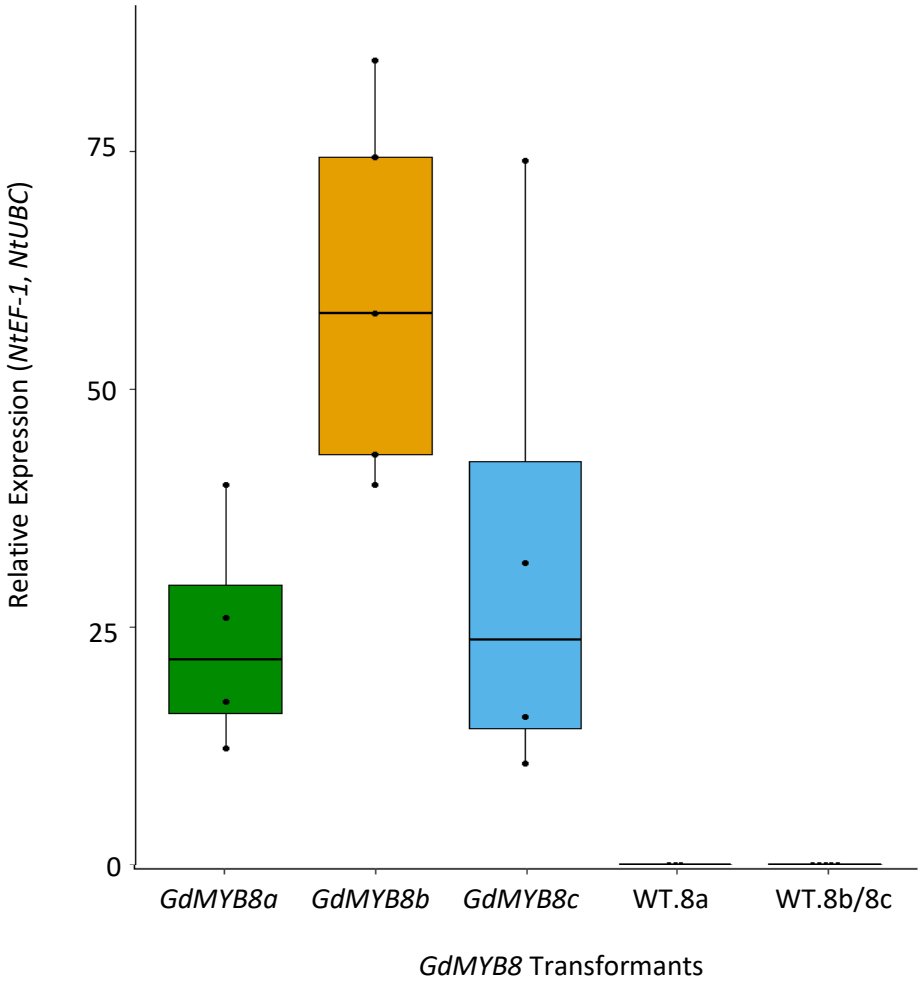


Figure 5.9. The relative expression of the transgene within several lines for each set of transformants. Primers were the same for *GdMYB8b* and *GdMYB8c* but differed for *GdMYB8a* – wild type expression levels are shown for each set of primers (WT.8a and WT.8b/8c). The black line in each box indicates the median value and the whiskers 25/75% quantile +/- 1.5 *IQR, respectively. Each data point (black dots) is from an independent line. Sample size is n = 4 – 5, each taken from an independent transgenic line/ wild type plant.

G. diffusa ANS, *DFR*, and *MAT1* are here being investigated as potential downstream targets of GdMYB8 regulation. The expression levels of the *N. tabacum* homologues (*NtANS*, *NtDFR*, and *NtMAT1*) were quantified in transgenic *N. tabacum* plants to determine whether any differences in biochemical functioning between GdMYB8 proteins could be inferred through comparisons of different transformants (Fig 5.10). There was a significant positive correlation between *GdMYB8* expression and the expression of *NtANS* ($t = 7.55$, $p = 0.001$, $R^2 = 0.77$) and *NtDFR* ($t = 4.28$, $p < 0.001$, $R^2 = 0.50$) across constructs. There was no difference in *NtMAT* expression between wild type and transgenic plants. Whereas, *NtANS* and *NtDFR* were significantly upregulated compared to wild type across *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* transformants (Appendix 5). A linear mixed model was conducted to determine whether *NtANS* and *NtDFR* were expressed at lower levels in *GdMYB8a* lines, while accounting for the variability in enzyme expression resulting from transgene expression level. In *GdMYB8a* transgenic lines upregulation of *NtANS* was much lower than in *GdMYB8b* ($t = -4.76$ $p = 0.001$) and *GdMYB8c* ($t = -2.50$ $p = 0.040$) counterparts, with transgene expression level accounted for. *NtANS* was also higher in *GdMYB8b* lines compared to *GdMYB8c* lines in these analyses ($t = 2.23$ $p = 0.044$). Fig. 5.11 demonstrates this with a comparison between a subset of lines. It is evident in Fig 5.11 that *GdMYB8a* transgenic lines had lower *NtANS* expression levels that *GdMYB8b* and *GdMYB8c* lines with similar transgene expression levels and anthocyanin content (e.g. 8a3 vs 8b1 vs 8c3 and 8a1 vs. 8a4 vs. 8c2). As only 4 independent lines of *GdMYB8a* transformants were available, a second plant

from each line was grown and expression levels quantified, and this pattern of *NtANS* expression was consistent within lines. Generally, lines with higher transgene expression had greater petal anthocyanin content - with a few exceptions.

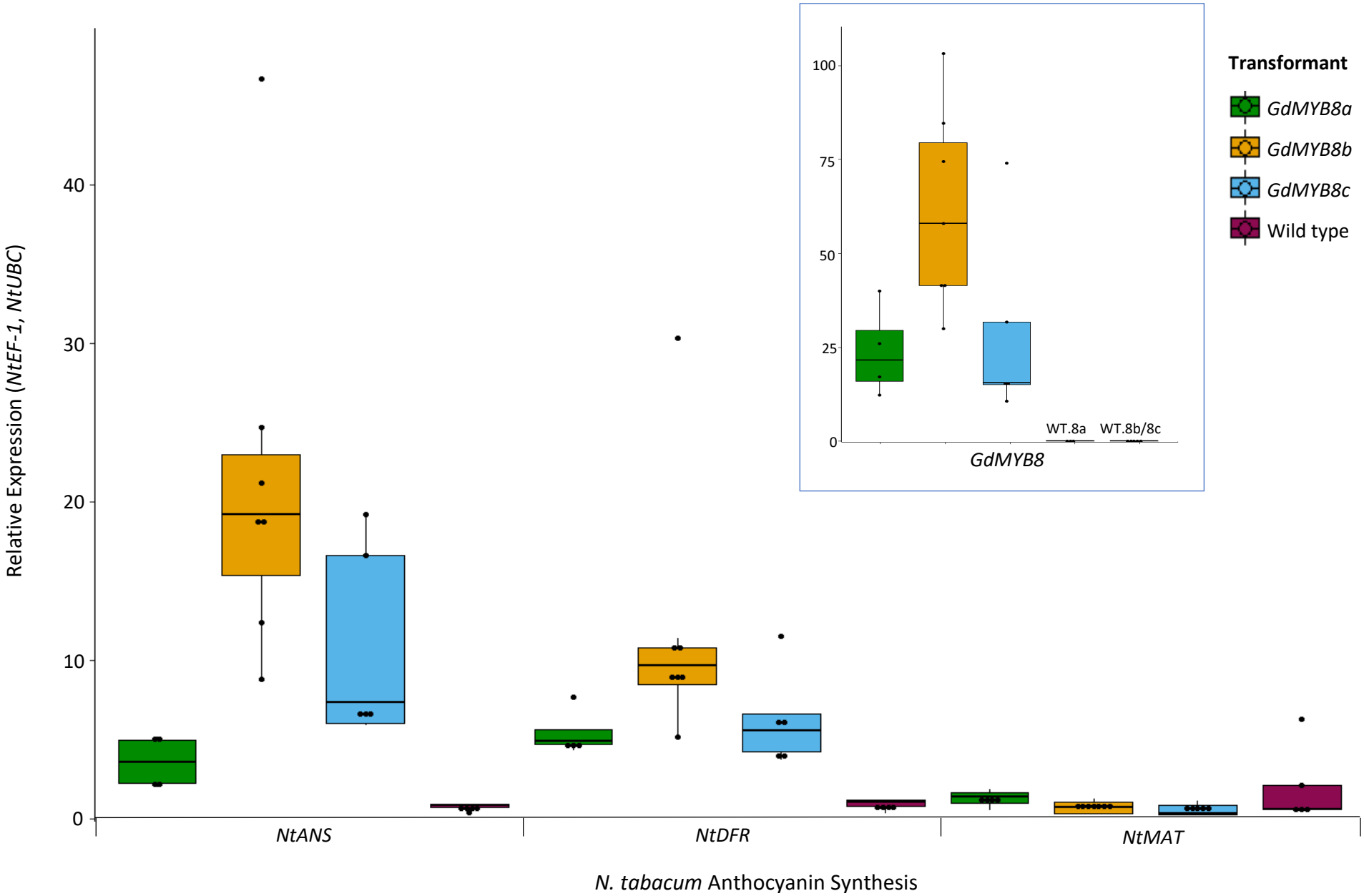


Figure 5.10. The relative expression of genes encoding *N. tabacum* anthocyanin synthesis enzymes (*anthocyanidin synthase* (*ANS*), *dihydroflavonol 4-reductase* (*DFR*), *malonyl transferase* (*MAT*)) in *N. tabacum* transformed with *G. diffusa GdMYB8* genes and wild type plants. The graph in the blue box is the relative expression levels of *GdMYB8* transgenes, presented in Fig. 5.9, with the addition of 3 samples for which pigment data was not available. The black line in each coloured box indicates the median value and the whiskers 25/75% quantile +/- 1.5 *IQR, respectively. Each data point (black dots) is from an independent line. Sample size is n = 4 – 7, each taken from an independent transgenic line/ wild type plant.

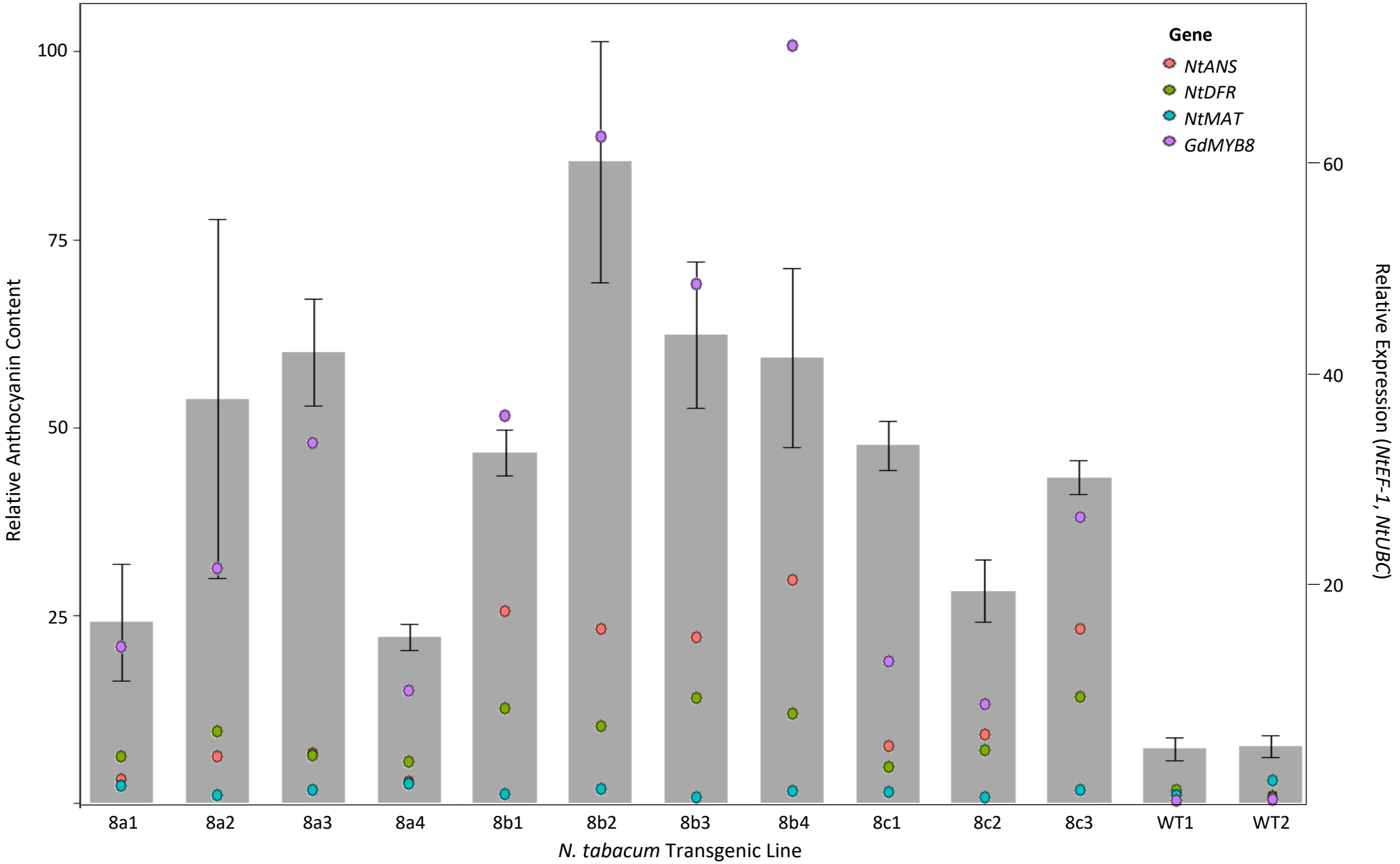


Figure 5.11. Comparison of the petal anthocyanin content and gene of interest expression levels (*NtANS*, *NtDFR*, *NtMAT*, *GdMYB8*) in individual plants from a subset of the independent *N. tabacum* transgenic lines. Individual lines are listed along the x axis with '8a' indicating an *N. tabacum* line transformed with *GdMYB8a* and the number corresponding to the line represented. Relative anthocyanin content is represented by the grey bars (mean ± s.e., n = 3) and the left-hand y axis. qRT-PCR expression levels are indicated by coloured dots, colour coded according to which gene they represent (see key) and expression levels are on the right-hand y-axis.

5.4 Discussion

In the previous chapter, three *GdMYB8* genes were identified as good candidates for encoding the regulators of *G. diffusa* petal spot anthocyanin pigmentation. Potential downstream targets of the GdMYB8 proteins were here investigated through characterisation of three genes encoding anthocyanin synthesis enzymes (GdANS, GdDFR, and GdMAT1). Expression analyses demonstrated that the genes encoding these enzymes were upregulated in developing petal spots compared to plain petal regions. This corresponds to spatial expression patterns of the *GdMYB8* genes, providing correlative evidence for regulation of the anthocyanin synthesis enzymes by GdMYB8 proteins. The capacity of GdMYB8 proteins to regulate anthocyanin synthesis was investigated through overexpression of each *GdMYB8* gene in *N. tabacum*. All GdMYB8 proteins activated anthocyanin production across a range of tissues in this heterologous system.

Multiple copies of *GdDFR* and *GdMAT* were identified, along with a single copy of *GdANS*. The number of gene copies for each enzyme could not be unequivocally determined as no genome sequence is available. Out of four *GdMAT1* variants present in an RNA-seq analysis (Kellenberger unpublished), one was upregulated in developing petal spots (*GdMAT1*) and so characterised here. At least three *GdDFR* copies were present, and they had highly conserved coding sequences, with more divergent introns. Other species have been shown to have multiple copies of *DFR*, for example, *Clarkia gracilis* (Martins et al. 2013). A few differences in amino acids occurred within the enzyme active site, which could potentially have functional consequences. A single or very few amino acid alterations to DFR were found to influence substrate specificity in several systems, including *Gerbera* and *Petunia* (Fischer et al. 2003; Johnson et al. 2001; Des Marais and Rausher 2008; Shimada et al. 2005). All *Petunia* and *Lotus japonicus* DFRs with an aspartic acid at a specific position were able to reduce the precursor for cyanidin synthesis, dihydroquercetin (DHQ), but not (or less readily for *L. japonicus*) dihydrokaempferol (DHK) (Johnson et al. 2001; Shimada et al. 2005). Consistent with these findings, all *GdDFR* variants have an aspartic acid at this position, and *G. diffusa* petals are pigmented solely by cyanidin. Premature stop codons in several *GdDFR* cDNA sequences indicate that truncated DFR proteins may be produced, containing only 27 - 62% of amino acids present in the full-length sequence. It is likely that shorter truncated proteins were non-functional, as key amino acids required for formation of the active site were missing (Johnson et al. 2001). It is also possible that alternative splicing of *GdDFR* occurs, with premature stop codons actually residing within introns. To determine if this is the case would require more extensive characterisation of the *GdDFR* cDNA produced by *G. diffusa* ray floret tissue. Functionality was more ambiguous for the longer sequences. While truncation of several anthocyanin synthesis proteins has been shown to result in loss of function (Deng et al. 2019; Hsu et al. 2017; Rafique et al. 2016; Rinaldo et al. 2015; Zufall and Rausher 2003), truncated MtMAT proteins of *Medicago truncatula* retained good catalytic efficiency, although optimal pH for maximum enzyme activity changed (Yu et al. 2008). To determine whether or not *GdDFR* truncated proteins are functional, assays investigating enzyme activity could be conducted. Given the presence of multiple *GdDFR* copies and its positioning within the anthocyanin pathway, further investigation of this enzyme could provide insight into mechanisms underlying spot pigmentation. Genome sequencing would enable conclusive identification of *GdDFR* genes and copy number.

Expression patterns of *GdMAT1*, *GdANS*, and *GdDFR* in the petals of Spring and Cal morphotypes suggest that these genes may be regulated by GdMYB8 proteins. All three enzyme-encoding genes were upregulated in developing petal spots compared to plain petal regions, consistent with *GdMYB8* spatial expression patterns. Petal tissue was collected at two developmental stages, the first when

petal spot development had just initiated, and during the second specialised spot cell types formed while abaxial petal ray floret pigmentation was produced. Abaxial anthocyanin pigmentation occurs in Spring plain ray floret petals, but not in spotted ray florets. In Cal petals, anthocyanin pigments upper abaxial regions, while basal regions contain little to no pigmentation. As such, expression levels of enzyme-encoding genes in spotted tissue at the second developmental stage are not the cumulative result of enzymes synthesising spot anthocyanin and abaxial anthocyanin. The overall expression levels of *GdANS* are greater than those of *GdDFR*. At the first developmental stage *GdANS* and *GdDFR* were significantly upregulated in spotted petal tissue only. Significant upregulation of *GdANS* and *GdDFR* occurred at the second developmental stage in both spotted and plain petal tissue. Upregulation in plain petal tissues indicates that these enzymes may contribute toward abaxial background pigmentation. As there was a single copy of *GdANS*, the increase in expression in spotted tissue between the first and second developmental stage was due to either continuous gene expression or a second wave of expression. *GdANS* expression levels did not differ between plain and spotted tissue at the second developmental stage in Spring or Cal, but *GdDFR* had higher expression levels in spotted tissue. Unfortunately, this is hard to interpret as *GdDFR* expression patterns are an amalgamation of those of several *GdDFR* genes. It is possible that the same *GdDFR* genes are upregulated to a greater extent in spotted regions or that there is a spot specific *GdDFR* allele or gene that is activated within the spot during the second developmental stage. In *Clarkia gracilis,* there is a spot specific *DFR* allele (*Dfr2*) that is activated prior to other *DFR* alleles and is expressed exclusively in the petal region where a spot will develop (Martins et al. 2013). Evidently, both *GdANS* and *GdDFR* may be important for spot anthocyanin production as they have spot specific expression at the first developmental stage followed by an increase in expression at the second. The latter is developmentally synchronised with the upregulation of *GdMYB8b*. There may also be additional anthocyanin synthesis enzymes important to spot anthocyanin production, that have greater divergence in expression levels between spotted and plain petal tissues.

There was significantly higher *GdMAT1* expression in developing petal spots compared to plain petal regions at both developmental stages investigated in Spring and Cal. Given that malonated anthocyanin is absent (or present in very small amounts) in plain petal tissue (Fig 4.7), these expression patterns are consistent with a role in catalysing the addition of malonyl residues to petal spot anthocyanins. This was further validated by no or trace *GdMAT1* expression levels found in plain Stein ray florets. *GdMAT1* expression levels were roughly equivalent to those of *GdMYB8a*, the candidate regulator with the lowest gene expression. No comparisons were found within the literature providing information on how the expression levels of malonyl transferases, shown to be functioning in anthocyanin synthesis, compare to those of other anthocyanin synthesis genes in other systems. A study in *Lilium* tepals compared expression between anthocyanin synthesis enzymes, reporting that some enzyme-encoding genes, such as *ANS* and *DFR,* had expression levels approximately 2.5-10 fold higher than the *MYB12* regulator, while others including *F3'H* and *3GT* (anthocyanidin 3-O glucosyltransferase) had expression levels equivalent to *MYB12* (Yamagishi 2018). Not all anthocyanins within the spot contain malonyl residues and lower expression of *GdMAT1* compared to *GdANS* and *GdDFR* is consistent with these findings. Ideally, functioning of *GdMAT1* would be determined by downregulating the gene within *G. diffusa*. This would enable an assessment of whether it is necessary for malonated anthocyanin production or if there is another gene with a higher expression level that has not yet been characterised.

All GdMYB8 proteins were capable of activating anthocyanin production in the heterologous host *N. tabacum*. Increased anthocyanin production was observed in all lines of *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* transformants. This demonstrated that GdMYB8 proteins were capable of partnering with *N. tabacum* co-regulators to regulate gene transcription. Anthocyanin pigment accumulated in multiple tissues including petals, anthers, sepals, the stigma, ovary walls, fruit, and leaves. All transformed lines had increased anthocyanin pigmentation in the petals compared to wild type, but not all had ectopic pigmentation in other tissues. This variation was likely due to differences in transgene expression level. Overall *GdMYB8b* lines had stronger phenotypes than *GdMYB8a* and *GdMYB8c* lines, particularly in sepals and leaves, and *GdMYB8b* transformants also generally had higher transgene expression levels. As such, strong *GdMYB8b* transformant phenotypes could either result from differences in transgene expression level or differences in functioning between GdMYB8 proteins within *N. tabacum*; these two possibilities are not mutually exclusive. Previous studies expressing subgroup 6 R2R3 MYB genes within *N. tabacum* have found similar phenotypes, including *GhMYB10* from the Asteraceae species *Gerbera hybrida* (An et al. 2015; Elomaa et al. 2003; Espley et al. 2007; Mooney et al. 1995; Tian et al. 2017). There was a significant positive correlation between transgene expression level and the expression levels of *NtANS* and *NtDFR* across constructs, suggesting that GdMYB8 proteins can co-ordinately control *NtANS* and *NtDFR*. GdMYB8 proteins did not regulate *NtMAT* as there was no difference in expression of *NtMAT* between transformants and wild type plants. In all *GdMYB8b* and *GdMYB8c* transformed lines, *NtANS* expression was higher than *NtDFR* expression. In contrast, *NtDFR* expression in *GdMYB8a* transformants was always greater than or equal to *NtANS* expression. *NtANS* was found to be expressed at significantly higher levels in *GdMYB8b* and *GdMYB8c* transformants compared with *GdMYB8a* lines, when the proportion of variation in *NtANS* expression explained by transgene expression level was accounted for. There was no obvious reduction in petal pigmentation as a result of this lower transgene expression, suggesting that *GdMYB8a* is likely to regulate additional *N. tabacum* anthocyanin synthesis enzymes. This does not exclude the possibility that *GdMYB8b* and *GdMYB8c* also regulate other *N. tabacum* anthocyanin synthesis genes.

In conclusion, the genes encoding these three anthocyanin synthesis enzymes (*GdANS*, *GdDFR*, and *GdMAT1*) are good candidates for regulation by GdMYB8 proteins. While *GdMAT1* may have a role exclusively in spot pigmentation, *GdANS* appears to be involved in pigmentation both in spots and in other regions of the ray floret petals. Temporal regulation of *GdANS* differ between regions, with gene expression activated within developing petal spots at an earlier developmental stage. *GdDFR* may have similar expression patterns to *GdANS* but due to the presence of multiple copies, that haven't been fully characterised, this cannot be confirmed. Heterologous expression of the *GdMYB8* genes indicate that they are all capable of activating anthocyanin production within *N. tabacum*. Differences in *NtANS* expression, between plants transformed with different *GdMYB8* genes, suggests that there may be differences in GdMYB8 protein biochemical properties that can lead to differential regulation of anthocyanin synthesis enzymes. However, whether differential regulation actually occurs within *G. diffusa* or whether it is a result of factors specific to *N. tabacum* systems remains to be determined. *GdMYB8b* and *GdMYB8c* were hypothesised to be the dominant regulators of petal spot anthocyanin due to high expression levels compared to *GdMYB8a*. Interestingly, *GdANS* is expressed at significantly higher levels than *GdDFR* across *G. diffusa* petal tissue – mirroring the pattern seen in *N. tabacum* plants expressing either *GdMYB8b* or *GdMYB8c*.

Chapter 6. Functional analyses of GdMYB8 in *Gorteria diffusa*

6.1 Introduction

'A key challenge in recent years has been to establish many of the experimental advantages found in model organisms in other species with divergent morphologies' (Monniaux and Hay 2016). Establishing robust molecular genetic techniques in non-model systems can provide insight into how complex phenotypes develop, within systems that enable coupling of this molecular mechanistic understanding with that of ecological and evolutionary knowledge. This interdisciplinary approach can be used to explore how diverse morphologies evolve (Monniaux and Hay 2016; Song and Mitchell-Olds 2011). Research into the wildflower genus *Mimulus* (monkeyflowers) exemplifies this approach. Recognised as a model system for evolutionary and ecological studies (Twyford et al. 2015; Yuan 2019), genomic tools have been developed across several *Mimulus* species within the last decade (Ding and Yuan 2016; Yuan et al. 2013a; Yuan et al. 2013b). These resources have been used for comparative genetic analyses and to gain insight into developmental processes. In the sister species *M. lewisii* and *M. cardinalis*, for example, genes underlying differences in flower colour have been characterised, and regulatory understanding of the spatial patterning of floral pigmentation has been developed (Yuan et al. 2016; Yuan et al. 2013a). Functional genetic analyses of candidate loci are conducted to determine gene function, and stable plant transformation is one technique that enables rigorous analysis of genetic hypotheses. Stable transformation protocols are available for *M. aurantiacus* (Streisfeld et al. 2013), *M. guttatus* (Preston et al. 2014), and *M. lewisii* (Yuan et al. 2013a). This technique was used to characterise the first transcription factor identified that activates floral carotenoid biosynthesis (RCP1) in *M. lewisii*. Downregulation of RCP1 reduced flower carotenoid content, while overexpression in a *rcp1* mutant background restored the production of carotenoid (Sagawa et al. 2016). Evidently, perturbing gene expression within the host organism is a key aim to enable understanding of the functional basis of complex traits.

Agrobacteria transfer and integrate a segment of bacterial DNA into the plant nuclear genome from a tumor-inducing (Ti) plasmid resident within the *Agrobacterium* cell. Processing and transfer of the DNA (T-DNA) from the bacterium into the plant cell largely results from the action of *virulence* (*vir)* genes that also reside on the Ti plasmid (Garfinkel and Nester 1980; Hooykaas et al. 1984; Horsch et al. 1986; Kohli et al. 1999; Lundquist et al. 1984; Stachel and Zambryski 1986). Manipulation of this system has enabled *Agrobacterium*-mediated stable plant transformation, which involves the genetic alteration of host plant species through insertion of specific DNA, for example, a gene of interest. Several binary Ti vectors have been designed for this purpose, enabling efficient transformation of plant cells. These transformed cells undergo regeneration to produce a mature transgenic plant, achieved through controlled hormone exposure and culturing of plant cells under tightly controlled physical and chemical conditions. *Agrobacterium tumefaciens* has been used to successfully transform a broad range of species (Chu et al. 1997; Kim et al. 2008; Elomaa et al. 1993; van Wordragen et al. 1991), including Asteraceae species such as *Gerbera hybrida* (Elomaa et al. 1993; Nagaraju et al. 1998). Overexpression of the *G. hybrida* protein GMYB10 greatly increased anthocyanin pigmentation in several tissues, including enhanced production of pelargonidin in the flower (Laitinen et al. 2008). Mellers (2016) used *A. tumefaciens* transformation to try to develop a *G. diffusa* stable transformation protocol using the Spring morphotype. These preliminary trials resulted in successful transformation of *G. diffusa* leaf discs and subsequent regeneration of transgenic calli into mature plants. These mature transformants expressed a transgene that produced a fluorescent protein, which was detectable in mature ray florets and leaves of the T_0 generation. While these preliminary

transformation results were promising, much optimisation remains including shortening of the regeneration time. One potential optimisation strategy, trialled here, is to alter the hormone concentrations during tissue culture as this may stimulate faster shoot and/or root growth.

Growing the T_1 generation of transgenic plants is a requirement for reliable phenotypic inference of transgene function, since the artificial hormonal environment of tissue culture can induce phenotypic aberration in the T_0 plants. It is of particular importance for *G. diffusa* transgenics given that T_0 floral phenotypes differed markedly from those of wild type plants, with petal spots frequently completely absent from capitula. When present, petal spot phenotype also differed from that of wild type spots regarding the cell types present and their positioning. The phenotype of spotted ray florets was also altered in transgenics, more closely resembling wild type plain ray florets than the shortened and raised ray floret phenotype characteristic of wild type Spring. As such, obtaining the next generation of transgenics is necessary to determine whether phenotypic alterations are due to stress induced by the regeneration process or if the observed impact on petal spot development is transgenerational. The latter would 'not only bring into question the usefulness of transformation to analyse gene function, but also question the underlying plasticity of spot character' (Mellers 2016). Previous attempts to cross the T_0 plants resulting from the Mellers (2016) stable transformation trial have failed; crossing was conducted rather than selfing because *G. diffusa* is self-incompatible (Ellis unpublished data). Additionally, wild type *G. diffusa* have never been successfully crossed within our laboratory. However, crossing has been achieved in glasshouse conditions in South Africa by another research group and so here attempts were made to obtain progeny of the Mellers (2016) transgenic *G. diffusa* plants. Analysing the floral traits of transgenic plant progeny would allow informed assessment of whether transformation and regeneration is a good tool for functional analysis of *G. diffusa* proteins.

Successful *G. diffusa* stable transformation would enable functional characterisation of GdMYB8 proteins. As such, attempted optimisation of the stable transformation protocol was conducted using a *GdMYB8* gene on a constitutive promoter, to see if overexpression of anthocyanin could be induced. The secondary aim of further developing the *G. diffusa* stable transformation protocol, was to enable production of transgenic plants expressing *GdMYB8* under its native promoter to determine whether spot production could be induced in a non-spotted *Gorteria*. While a stable transformation protocol is favourable for these more complex regulatory questions, to determine whether GdMYB8 proteins can regulate anthocyanin production, as predicted, transient expression in *G. diffusa* would suffice. This approach has been used in several studies investigating genes hypothesised to be involved in anthocyanin synthesis or regulation across multiple systems including *Mimulus* (Ding and Yuan 2016), *Gerbera jemosonii* (Hussein et al. 2013), *Antirrhinum majus* (Shang et al. 2011) and *Clarkia gracillis* (Martins et al. 2017). A preliminary *G. diffusa* transient transformation trial was attempted here.

Previous chapters have provided correlative evidence demonstrating that GdMYB8 proteins may regulate *GdANS*, *GdDFR*, and *GdMAT1*. Given the complexity, unpredictability, and time taken to develop transformation protocols in non-model systems, an alternative approach was adopted in parallel to provide additional evidence for whether or not GdMYB8 proteins function in anthocyanin regulation. Promoter regions of *GdANS*, *GdDFR*, and *GdMAT1* were characterised and the ability of GdMYB8 proteins to bind to these regions was tested through yeast one-hybrid experiments and gel shift assays (i.e. electromobility shift assays). While not as conclusive as *G. diffusa in vivo* experiments, consistent positive results between these two complementary techniques would provide a good

indication that proteins are capable of interacting with regulatory elements of the genes they are hypothesised to regulate. Regulation of genes encoding anthocyanin synthesis enzymes by R2R3 MYBs is well documented across a number of species. Within several systems yeast one-hybrid experiments and/ or gel shift assays have been used to confirm interactions between promoter regions of the genes encoding anthocyanin synthesis enzymes and MYB regulators (Kelemen et al. 2015; Tian et al. 2017; Wang et al. 2018; Yong et al. 2019). Resources are also available detailing predicted conserved binding motifs for subgroup 6 R2R3 MYB transcription factors (e.g. O'Malley et al. 2016).

In combination with previous research presented within this thesis, determining functionality of GdMYB8 proteins would mark the first comprehensive investigation into the regulation of a spot-specific genetic pathway in *G. diffusa*. Here attempts were made to further develop the *G. diffusa* transformation protocol, assessing whether regeneration time of stable transformants could be reduced. Crossing of transgenic T_0 plants was attempted to obtain T_1 progeny. Preliminary trials of transient transformation within *G. diffusa* were conducted to assess its feasibility in this system. The above methods were used in combination with more established methods (yeast one-hybrid experiments and gel shift assays) for non-model systems to determine whether GdMYB8 proteins were able to bind to promoter regions of genes encoding *G. diffusa* anthocyanin synthesis enzymes and activate gene expression.

6.2 Methods

6.2.1 *Gorteria diffusa* stable transformation

This protocol is a slightly modified version of that presented in Mellers (2016). *Agrobacterium tumefaciens* strain GV3101 was used to transform the Spring morphotype of *G .diffusa* to produce plants constitutively expressing the GdMYB8a protein through insertion of a pGreen construct containing *GdMYB8a* under the control of a double 35S CaMV promoter and a 35S terminator sequence (vector diagram in Fig 5.1). The transformation of *A. tumefaciens* with the vector is outlined in Sections 2.6.2.

All procedures detailed were conducted in sterile conditions under a laminar flow hood (except for leaf removal from plants). Preparation of *A. tumefaciens* solution is described in Section 2.6.3, with the alteration that pellets were resuspended in MS9 (Appendix 6) and 50µl of freshly prepared 100mM acetosyringone solution were added. Following resuspension through gentle mixing the solutions were shaken at 28°C for at least 30 mins before being used to infect leaf tissue. Mature *G. diffusa* leaves were removed from the plant and sterilised in 10% bleach and 0.1% (w/v) SDS solution for 20 mins, rinsed several times in sterile ddH_2O, and the water was immediately drained off. A petri dish was filled with the *Agrobacterium* suspension and multiple leaves were submerged in the solution. The edges of each submerged leaf were removed, and discarded, and the remaining leaf was cut latitudinally into 1cm^2 segments to maximise the surface area of the leaf susceptible to *Agrobacterium* infiltration. To remove excess *Agrobacterium*, the leaf squares were transferred to filter paper and then placed adaxial side up on a co-cultivation media (CCM) plate (Appendix 6). The CCM plates containing leaf discs were then incubated for 48 hrs in the dark at room temperature to promote *Agrobacterium* infection.

The hormone concentrations within shoot inducing medium were altered from previous transformation attempts (Mellers 2016). Explants were grown on SIM containing a two-fold increase in cytokinin (trans-Zeatin) concentration (2mg/l) and a reduced auxin (IAA) concentration (1mg/l), as compared to *G. diffusa* transformations performed previously. These alterations were made because trans-Zeatin has been shown to induce the growth of shoot buds from explants within other systems (Coleman and Ernst 1989; Ćosić et al. 2015) and auxin was found to be necessary for regeneration in a previous transformation trial (Mellers 2016). After 48 hrs in the dark, leaf discs were transferred to new complex shoot inducing media (cSIM) plates (Appendix 6). For the remainder of the tissue culture period plates were kept in a 16 hrs light - 8 hrs dark regime at 24°C with plant tissue moved to fresh plates every 7 - 10 days. After the initial 7 days on cSIM plates, all subsequent plates were made of regular shoot inducing media (rSIM) (Appendix 6). The addition of kanamycin to the medium selected against plant cells that did not contain the pGREEN plasmid and cefotaxime prevents agrobacteria growth. Once plantlets differentiating from calli had reached 1 - 2cm in size these were carefully prised off with tweezers and transferred to 50ml Hamilton jars containing SIM (Appendix 6) and either cefotaxime, ampicillin or no antibiotics to determine whether rooting was inhibited by these antibiotics. Once sufficient root stock had formed (two well developed roots reaching the base/ sides of the Hamilton jar) plantlets were transferred to plant pots and grown in an 80:20 ratio of Levington's M3 bedding compost and sand in a glasshouse (16 hrs light/ 8 hrs dark, 20 °C (daytime), 60% humidity, fluorescence of 200 µmol/m/s). Cuttings were taken by severing a healthy young stem from the parent plant, dipping it in water and then rooting powder (1-Naphthaleneacetic acid, Bayer Strike hormone rooting powder), and planting it in soil.

To determine whether the transgene was being expressed, leaf RNA was extracted, cDNA was synthesised, and PCR was conducted to amplify the transgene. PCRs were also run using RNA templates to check for gDNA contamination, indicated by the presence of a band (primers in Appendix 2). Leaves were used because they do not express *GdMYB8a,* enabling a comparison between wild type and transgenics leaves to confirm whether plants were transformed (Fig 6.1).

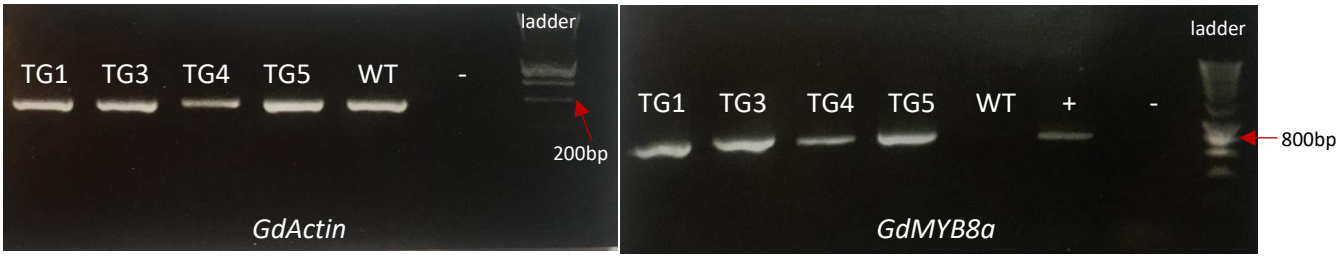


Figure 6.1. cDNA amplification of the transgene *GdMYB8a* in 4 out of 5 transgenic plants (TG1, 3, 4, 5). The *GdActin* PCR on the left-hand side demonstrates that cDNA synthesis was successful for transgenic and wild type (WT) samples (expected length 202bp). The *GdMYB8a* PCR amplified the transgene (expected length 795bp) in all plants except the wild type. The negative control contains no template and the positive control is wild type *G. diffusa* floral bud tissue.

6.2.2 Crossing *G. diffusa* plants

G. diffusa is self-incompatible (Ellis unpublished) and so crossing of transformed T_0 *G. diffusa* plants was attempted. A disc floret containing a pollen presenter visibly covered in pollen was plucked from the capitulum of one plant using tweezers. A disc floret with a receptive stigma, defined by stigma developmental stage (illustrated in Fig 6.2), was identified in the capitulum of another plant. The plucked pollen presenter was gently wiped up and down on the surface of the receptive stigma. This process was repeated for approximately 2 additional disc florets within the same capitulum, the capitulum was then tagged with a label indicating the paternal identity. Several capitula were hand pollinated per plant. Successful germination is indicated by lignification of the infructescence after anthesis. Crossing was trialled between transgenic plants ($n_{plants} = 5$), between transgenic and wild type plants ($n_{plants} = 5$), and between wild type plants ($n_{plants} = 8$). Crossing was attempted using between 5 - 10 plants approximately once a week over a 4 month periods, depending on the availability of pollen, open capitula, and the developmental stage of disc florets.

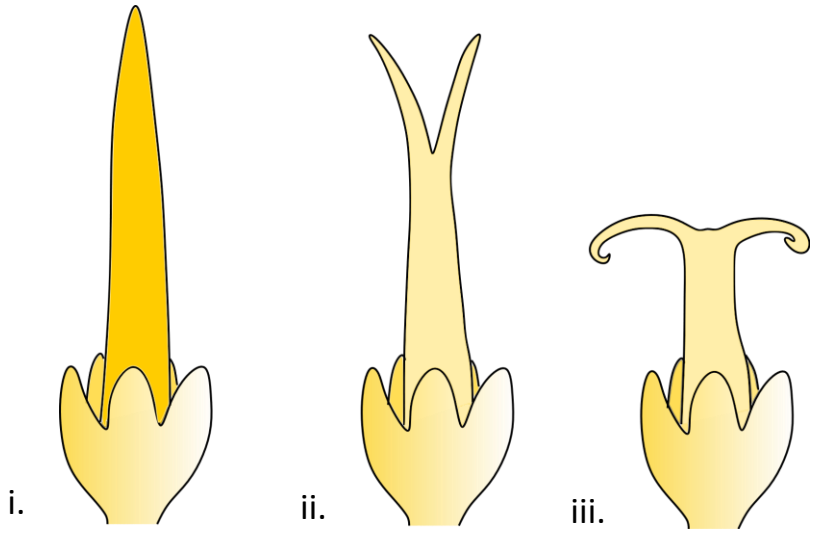


Figure 6.2. Schematic of *G. diffusa* disc florets at various developmental stages used for crossing attempts. i) illustrates a disc floret containing a pollen presenter covered in pollen. These disc florets were plucked from the capitulum and the pollen was wiped onto the receptive stigma of a disc floret depicted in iii). Subsequently, we were advised by a collaborator that crossing attempts are more successful when conducted using stigma at the developmental stage in ii).

6.2.3 HPLC analysis of regenerated plantlets

Samples were prepared as outlined in Section 2.8. Acidic methanol extractions were then analysed by high performance liquid chromatography (HPLC; Surveyor system, Thermo Scientific, San Jose, CA, USA) using photodiode array spectrometry. Data were analysed using Xcalibur software (Thermo Fisher Scientific). A 20µl injection volume of each sample was resolved on a Luna C18 column using 0.5% (v/v) formic acid (solvent A) and acetonitrile (solvent B) with a 0.2ml/min flow rate and an increasing gradient of solvent B (Appendix 6). The diode array detector collected spectra from 200-650nm. The identification of metabolites was based on their absorbance spectra compared to a reference cyanidin 3-O-glucoside (Sigma).

6.2.4 *Gorteria diffusa* transient transformation

A. tumefaciens GV3101 containing a pGREEN-GFP vector were grown from a glycerol stock overnight in LB (Appendix 1) with kanamycin (50mg/l), gentamycin (25mg/l), and rifampicin (50mg/l) at 28°C and stirred at 130rpm. Once the OD_{600} reached 1.0 - 2.0 the solution was centrifuged for 4 mins at 5000xg, the supernatant was removed, and the pellet was resuspended in infiltration media (for 100ml: 10ml 100 mM $MgCl_2$, 200µl 50mM acetosyringone, 2ml 500mM MES (pH 5.6), 87.8ml ddH_2O). Mature Spring *G. diffusa* plants were used, and different ages of leaf were trialled for infiltration. Healthy leaves were selected and, while still attached to the plant, a small cut was made through the adaxial surfaces of the leaf/ fused ray floret petals with a razor blade. *A. tumefaciens* solution was inserted into the cut using a syringe. A marker pen was then used to draw around the area which had been infiltrated. Plants were watered on the soil every 1 - 2 days and transgene expression was checked each day using a GFP filter under the microscope. Wild type capitula and wild type leaves of infiltrated plants were used for comparison.

6.2.5 Isolating upstream regions of *GdMYB8* genes and genes encoding anthocyanin synthesis enzymes

Genome walking (Section 2.3.3) was used to sequence upstream regions of *GdANS*, *GdDFR*, *GdMAT1*, *GdMYB8a*, *GdMYB8b* and *GdMYB8c*. All primer sequences are in Appendix 2.

GdANS, *GdDFR*, and *GdMAT1* promoters were being sequenced for use in experiments to see whether GdMYB8 proteins were capable of binding to motifs in these sequences and activating transcription. All genome walking for genes encoding anthocyanin synthesis enzyme was conducted using the Universal GenomeWalker 2.0 kit. Reverse primers were initially designed to amplify within target genes and, subsequently, within sequences upstream of the gene to isolate regions further from the start codon. *GdMAT1* primers were designed to be specific to this gene and several *GdDFR* primers were used to try to capture the diversity of the different variants and amplify upstream of all *GdDFR* variants. The full length of the promoter region found, and the corresponding gene, were amplified in a single PCR reaction. The PCR product was sequenced to ensure that the upstream region was correct. Raw *G. diffusa* genome sequencing reads became available very recently with an approximately 15% error rate. 1.5kb upstream of *GdDFR* had already been isolated, but these genomic reads were used to design forward primers to amplify further upstream of *GdANS* and *GdMAT1*.

The upstream regions of *GdMYB8a*, *GdMYB8b* and *GdMYB8c* were isolated using genome walking without a kit. As the genome was cut into fragments using restriction enzymes and circularised through ligation, when additional sequences were isolated it was unknown whether they were 5' or 3' of the gene. Therefore, for every additional sequence found, several forward primers were designed along its length, with reverse primers within the gene, and additional PCRs were conducted to determine if upstream regions had been isolated. The kit was subsequently used once no more progress could be made with this method.

6.2.6 Yeast one-hybrid (Y1H) experiments

6.2.6.1 Basic concepts

Yeast one-hybrid (Y1H) experiments were conducted to determine whether the GdMYB8 proteins were capable of binding to promoter regions of *GdANS*, *GdDFR*, and *GdMAT1* genes and activating transcription within yeast cells. In Y1H analysis, the interaction between a transcription factor ('prey': GdMYB8a, GdMYB8b, or GdMYB8c) and a DNA sequence upstream of a reporter gene ('bait': *GdANS*, *GdDFR*, or *GdMAT1* promoter fragments) is detected *in vivo* in yeast. The experiment was conducted in the yeast strain PJ69-2A (*MATa*, *trp1-901*, *leu2-3, 112, ura3-52, his3-200, gal4Δ*, *gal80Δ*, *LYS2 : : GAL1*$_{UAS}$-GAL1$_{TATA}$-HIS3, James et al. 1996) that is auxotrophic for leucine (*leu2-3*) and histidine (*his3-200*). A functional copy of the *LEU2* gene was used as a positive selection marker for prey plasmid, where a GdMYB8 protein was fused to the Gal4 activation domain (GAL4-AD) to ensure that binding of the GdMYB8 protein to the bait would result in reporter gene activation. The bait plasmid contains *HIS3* downstream of the bait DNA sequence as a reporter gene for prey-bait interactions (Figure 6.3). See further explanation below.

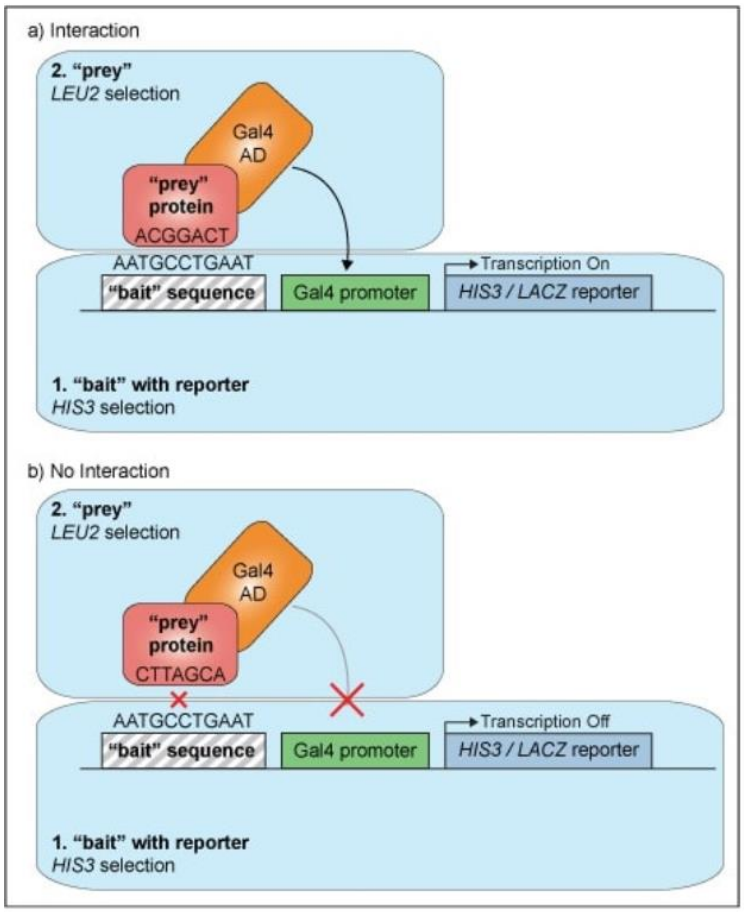


Figure 6.3 Schematic of how Y1H works. Upon binding of the prey protein to the bait sequence, the Gal4-AD interacts with the Gal4 promoter activating transcription of the *HIS3* reporter. Image taken from van Geldermalsen (2016).

6.2.6.2 Constructing and preparing plasmids

Bait plasmids

The upstream region of each gene encoding an anthocyanin synthesis enzyme was divided into overlapping regions of between 162 - 200bp. Each of these fragments was inserted upstream of the reporter gene *HIS3* in the pHISi vector (Fig. 6.4a). This was achieved through digestion of PCR fragments and pHISi with a pair of restriction enzymes (EcoRI and SacI) and subsequent ligation (described in Section 2.5.2). Vectors were transformed into *E. coli*, transformed cells were cultured, and plasmids purified using a miniprep protocol (Sections 2.4.2 and 2.4.3). Bait plasmids were linearized using the restriction enzymes XhoI or AflII (when additional XhoI sites were present in the plasmid) that cut within the *HIS3* gene in order to facilitate genomic integration into the *his3-200* locus. 1µg of purified plasmid was digested in a 37^{o}C incubator over 2 hrs using: 2µl enzyme, 6µl CutSmart buffer, 1µg template, and sterile DI water in a 40µl reaction.

Prey plasmids

The coding sequences of *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* were inserted in frame with the GAL4-AD into the pC-ACT2 vector (Fig 6.3b) using gateway cloning (Section 2.5.3). Plasmids were purified as explained above for pHISi vectors and 100ng of *GdMYB8* pC-ACT2 vectors were used for yeast transformation. All subsequent steps were done with sterile equipment in a laminar flow hood.

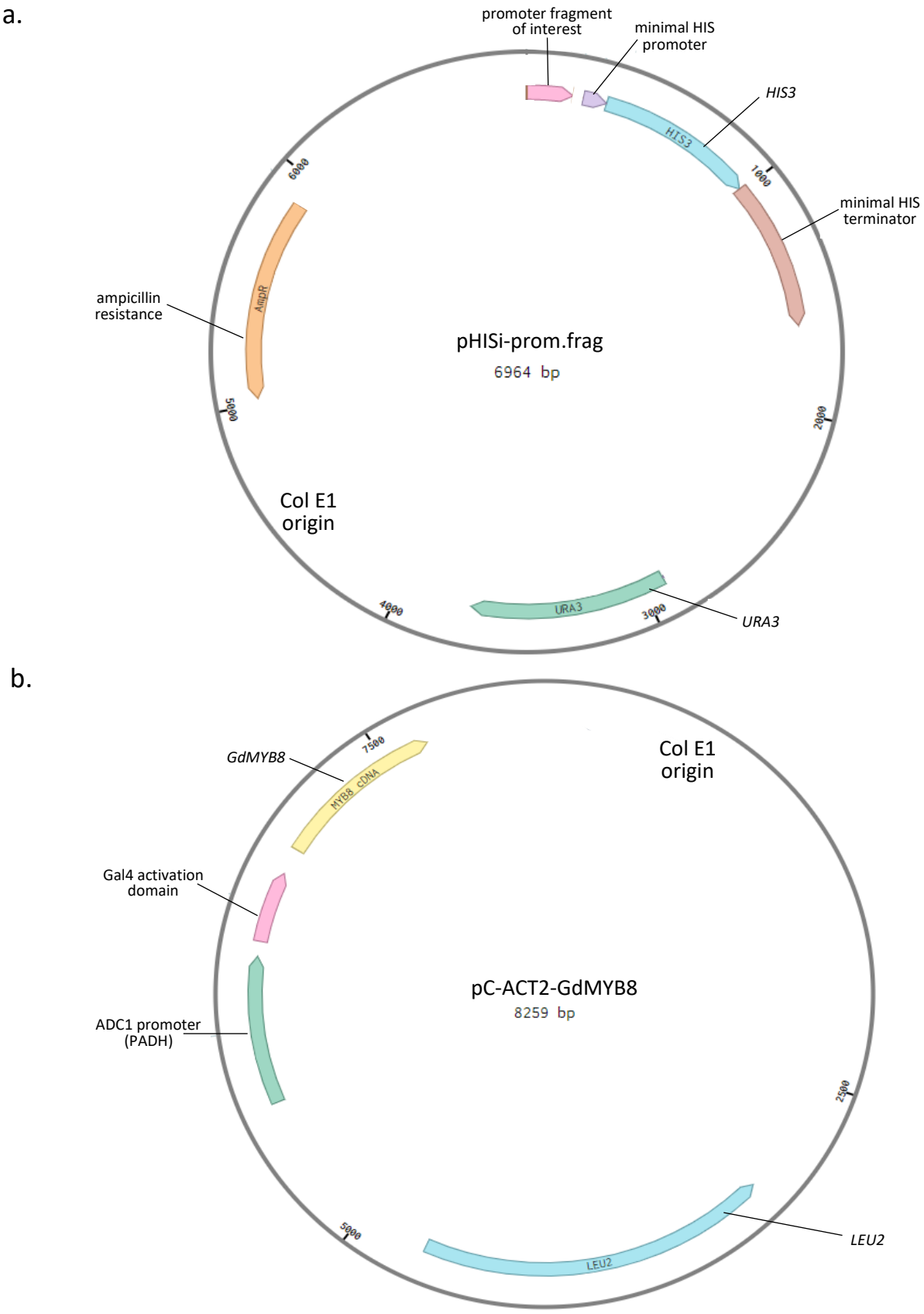


Figure 6.4. Vectors used in Y1H experiments. a) pHISi vector containing a fragment from upstream of a gene encoding an anthocyanin synthesis enzyme (*GdANS*, *GdDFR*, or *GdMAT1*), 'the bait'. This vector was linearised and enters the yeast genome through recombination. b) pC-ACT2 vector containing either GdMYB8a, GdMYB8b, or GdMYB8c fused to the GAL4 activation domain (GAL4-AD), 'the prey'. If the GdMYB8 protein can bind to a promoter fragment, the GAL4 activation domain ensures that transcription of the reporter gene is activated, in this case *HIS3.*

PJ69-2A yeast was streaked onto a YPDA plate (Appendix 6) and incubated at 30°C for 2 - 3 days. 3 colonies were then picked and transferred into 10ml of YPD medium in a 50ml falcon tube in sterile conditions. The suspension was vortexed and topped up to 50ml with YPD media, this was shaken at 180rpm at 30°C for 16 - 18 hrs. The density of the culture was measured with a spectrophotometer. The volume of overnight culture required to produce an OD_{600} = 0.2 - 0.3 in a larger (300ml) culture was transferred into a conical flask containing 300ml of YPD media. This was incubated at 30°C for 2 - 3 hrs with shaking at 230rpm and, once an OD_{600} of 0.5 +/- 0.1 was reached, the culture was centrifuged at 1000xg for 5 mins at room temperature. The supernatant was removed, and 25ml of 1xTE/1xLiAc (5ml 10x TE, 5ml 10x LiAc, 40ml sterile ddH_2O) were added and the pellet resuspended by gentle mixing. Again, this was centrifuged at 1000xg for 5 mins at room temperature. The supernatant was discarded, and the pellet resuspended in 1.5ml of freshly prepared sterile 1xTE/1xLiAc producing a solution of yeast competent cells.

Carrier DNA (UltraPure Salmon Sperm DNA Solution, Thermofisher) was linearised by heating at 98°C for 10 mins and then cooled on ice. 10µl of linearized Carrier DNA was added to 1µg of linearised pHISi plasmid in a microcentrifuge tube and mixed well by pipetting. 100µl of the yeast competent cells were added to the plasmid and mixed by vortexing. 600µl of sterile PEG/LiAc (1ml 10x TE, 1ml 10x LiAc, 8ml 50% (w/v) PEG) were added, the solution was vortexed and then incubated at 30°C for 30 mins with shaking at 200rpm. 70µl of DMSO were added and mixed in gently by inversion, the solution was heat shocked for 15 mins at 42°C and chilled on ice for 2 mins. Following centrifugation for 5 secs at 10000xg the supernatant was removed, and the cells resuspended in 500µl sterile 1x TE buffer. The cells were plated on synthetic dropout lacking histidine (SD-HIS; for 1L: 100ml 10x -HIS dropout solution (Appendix 6), 6.7g nitrogen base w/o amino acids, 20g agar, 2% (v/v) glucose) using a sterile spreader and left incubating at 30°C for 2 - 3 days. Successful genomic integration of the linearized bait plasmid into the *his3-200* locus results in leaky expression of *HIS3,* allowing yeast to grow on SD -HIS plates. An empty pHISi vector was transformed into yeast as a negative control. A non-linearised plasmid was also 'transformed' into yeast as a negative control and no yeast grew on these plates. Four colonies per transformation were selected at random and streaked onto fresh SD -His plates and incubated at 30°C for 2 - 3 days to provide multiple clones for subsequent transformations.

6.2.6.4 Yeast colony PCR

To confirm the insertion of the prey into the yeast genome, a colony PCR assay was used. First, yeast genomic DNA was extracted as follows. A single yeast colony was picked from the plate and resuspended in 100µl of 200mM LiAc and 1% SDS solution. This was incubated for 5 mins at 70°C and 300µl of 100% (v/v) ethanol added. The solution was vortexed and centrifuged at 10000xg for 3 mins, the supernatant was removed. The pellet was then washed with 500µl of 70% (v/v) ethanol and centrifuged at 10000xg for 1 min. The supernatant was removed, and the pellet resuspended in 100µl of 1x TE buffer. The solution was centrifuged again at 10000xg to pellet cell debris and 1µl of the supernatant containing genomic DNA was used in PCR (Section 2.2.3) and run on an agarose gel to check for amplification.

6.2.6.5 Spot tests for leaky HIS3 expression

The *HIS3* reporter gene has leaky expression also known as background activity, where recognition of the bait sequence by endogenous yeast transcription factors leads to activation of *HIS3* that is sufficient for growth on media lacking histidine. 3-amino-1, 2, 4-triazole (3-AT) is a competitive inhibitor of the His3 enzyme and can be used to supress the growth caused by leaky *HIS3* expression

of yeast strains. To determine the minimal 3-AT concentration that inhibits the growth of each bait strain a spot test was conducted. For each bait strain 4 colonies were individually picked using a sterile cocktail stick and resuspended in 150µl of 1x TE, from this a 10-fold and 100-fold dilution was made. 3µl of each concentration were pipetted onto SD -HIS plates varying in the 3-AT concentration they contained (0, 5mM, 10mM, 20mM, 30mM, 40mM, 50mM, 60mM), once the spots of liquid were dry plates were kept at 30°C for 2 - 3 days. After which, photographs of the plates were taken and the colony with the greatest susceptibility to 3-AT was selected to be transformed with pC-ACT2 vectors containing the *GdMYB8* encoding sequences (Figure 6.4b).

6.2.6.6 Transforming HIS3 PJ69-2A yeast strains with GdMYB8 pC-ACT2

For each bait strain, the colony with the least background activity was selected and used for 5 transformations with the pC-ACT2 vector containing *GdMYB8a*, *GdMYB8b* and *GdMYB8c* fused to the GAL4-AD. Additionally, pC-ACT2 plasmids containing *RVE-1* and a gene encoding the fluorescent protein Venus fused to GAL4-AD were used as negative controls. GdMYB8 proteins and RVE-1 are from different clades of the MYB protein family, so the latter was used to demonstrate that GdMYB8 binding properties are not generic to all MYBs. The transformation procedure was as outlined in 6.2.4.3 except carrier DNA was added to 100ng of (non-linearised) pC-ACT2 vector and transformed colonies were plated on SD -HIS -LEU media using 10x -HIS -LEU dropout solution (Appendix 6). The spot test was carried out (as described in Section 6.2.6.5) on SD -HIS -LEU media containing increasing 3-AT concentrations starting from the minimal 3-AT concentration determined in Section 6.2.6.5. Each bait strain was independently transformed with the five pC-ACT2 vectors twice to ensure that binding patterns were consistent. Transformations of vectors into yeast transformed with an empty pHISi vector (i.e. containing no bait sequence) were also conducted as a negative control.

6.2.7 Gel shift assays (EMSA)

A gel shift assay was conducted to determine whether GdMYB8 proteins could bind to motifs isolated from upstream regions of anthocyanin synthesis enzymes. This would complement the one-hybrid experiments and provide more thorough evidence for potential regulation by GdMYB8 proteins. Unfortunately, due to COVID-related access restrictions to the Sainsbury Laboratory Cambridge University the experiment could not be completed.

6.2.7.1 E. coli transformation and protein induction

GdMYB8a, *GdMYB8b,* and *GdMYB8c* coding sequences were inserted into individual pETM-11 vectors (Fig 6.5) by Gibson assembly (Section 2.5.1) to produce a protein with GdMYB8 tagged by 6 histidines at either end. These plasmids were transformed into the rosetta II strain of *E. coli* that is able to enhance eukaryotic protein expression by supplying tRNAs for specific codons that are rarely used in *E. coli*. The transformation procedure is outlined in Section 2.4.3, with the modification that the antibiotic selection used for growth on LBA plates and LB medium (Appendix 1) were 30mg/l chloramphenicol and 50mg/l kanamycin. Successful transformation was confirmed through colony PCR (Section 2.4.3). A single colony was picked from each transformation and grown in LB with kanamycin and chloramphenicol and stirred at 180rpm for 16 -18 hrs at 37°C. 3ml of culture were then removed, added to 250ml of LB with 100mg/l kanamycin, and grown at 180rpm for 2 - 3 hrs at 37°C. Once the OD_{600} of each culture reached approximately 0.6, 250µl of IPTG and 500µl of 1M betaine were added to induce protein production and increase the level of soluble protein, respectively. Control cultures were also set up in parallel in which protein production was not induced. The cultures were incubated and stirred for a further 2.5 hrs. 3ml of each culture was used to make 3 aliquots per

culture in microcentrifuge tubes. The microcentrifuge tubes and the rest of the culture in 50ml falcon tubes were spun down at 5000rpm, the supernatant was removed, and the pellets containing the GdMYB8 proteins were frozen at -20°C.

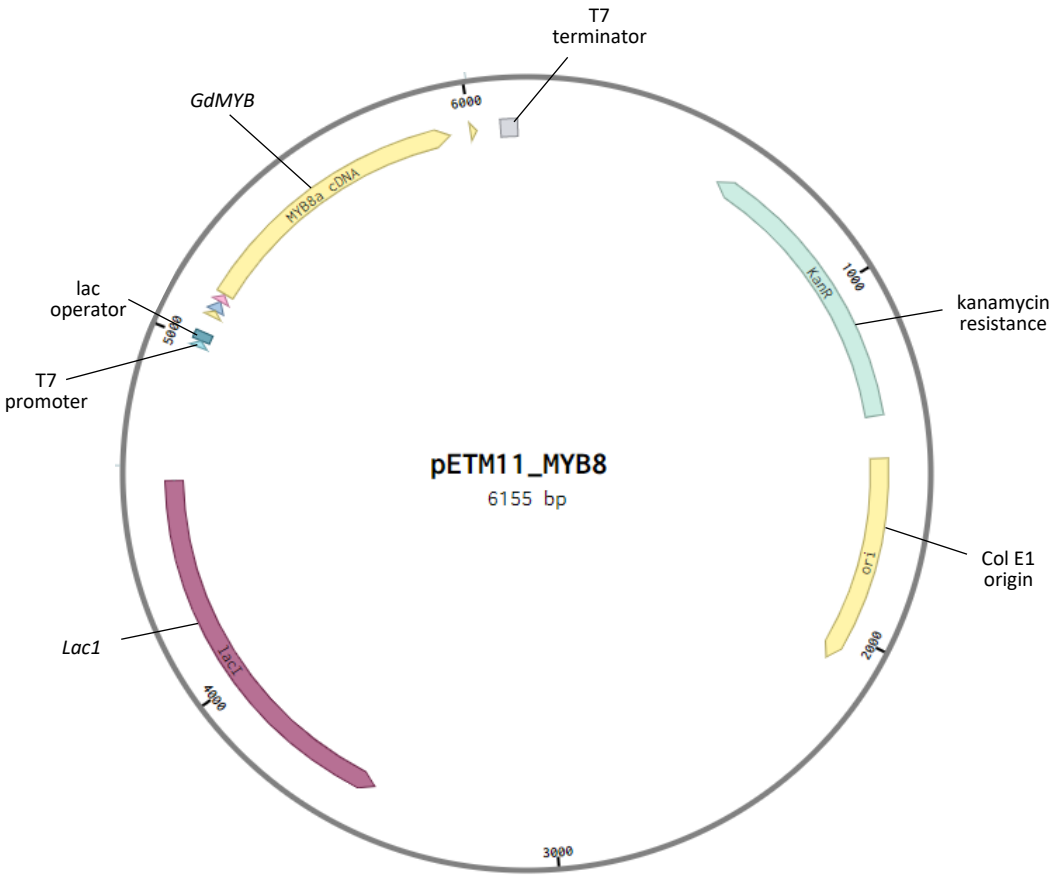


Figure 6.5. Diagram of the pETM11 vector used to induce production of histidine tagged GdMYB8 proteins in *E. coli* for use in gel shift assays. Three different vectors were constructed differing only in whether they contained *GdMYB8a*, *GdMYB8b*, or *GdMYB8c*.

6.2.7.2 SDS acrylamide protein gel

Microcentrifuge tubes containing pellets of *E. coli* transformed with the vectors containing the GdMYB8 proteins (induced) and control pellets (non-induced) were thawed and resuspended in 1ml of sterile ddH_2O. 20µl of this were mixed with 7µl of 4x NuPAGE LDS Sample buffer (thermofisher) and heated at 70°C for 10 mins to denature the protein. A discontinuous gel was produced from two acrylamide solutions. The stacking gel (Table 6.1), within which the protein wells are formed, enables protein stacking where proteins are concentrated in a thin zone. The resolving gel (Table 6.1) separates the proteins by size. The appropriate SDS acrylamide gel concentration was calculated using the expasy protein parameter tool (https://web.expasy.org/protparam/) and for the histidine tagged GdMYB8 proteins (37kDa) a 12.5% (v/v) resolving gel was recommended. The resolving gel solution was pipetted into the gel holder and 100% (v/v) ethanol poured on top to prevent it from drying out. Once this gel was set the ethanol was drained off and the stacking gel solution was added. Once set, the gel was immersed in a tank of Laemmli buffer (Appendix 6), samples of denatured proteins were loaded into the wells, and the gel was run at 120V for approximately 35 mins. Once the gel had run, it

was put in a sealed container submerged in Coomassie blue (Appendix 6) and left stirring overnight. The next day, the Coomassie blue was removed, and the gel was rinsed in ddH_2O and then submerged in destaining solution (50ml glacial acetic acid, 150ml ethanol, 300ml ddH_2O) until the majority of the blue stain was removed. Gel photographs were taken with a Nikon COOLPIX P520 camera (Fig 6.6).

Resolving gel		**Stacking gel**	
ddH_2O	3.2ml	ddH_2O	3.4ml
1.5M Tris-HCl pH 8.8	2.5ml	0.5M Tris-HCl pH 6.8	1.5ml
Acrylamide/bisacrylamide (37:5:1)	4.2ml	Acrylamide/bisacrylamide (37:5:1)	1.0ml
10% SDS	100µl	10% SDS	60µl
10% APS	50µl	10% APS	35µl
Temed	5µl	Temed	6µl

Table 6.1. SDS acrylamide recipes for 10ml of resolving gel and stacking gel used to determine whether protein induction in *E. coli* was successful.

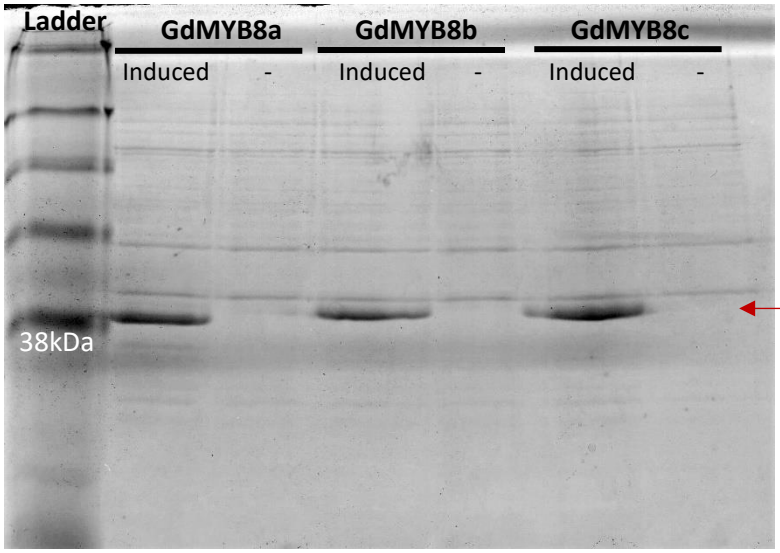


Figure 6.6. Acrylamide gel checking whether GdMYB8 protein production has been successfully induced in *E. coli*. Protein induction has been successful indicated by the band in induced cultures but lacking in the non-induced negative control cultures. The ladder is SeeBlue Plus2 Prestained Standard (novex). The bands of induced proteins are at the expected size of 37kDa.

6.2.7.3 Protein sonication and purification

After protein induction, the *E. coli* cells were sonicated to release cellular contents and enable testing of protein solubility. Base buffer was prepared (Appendix 6) and the pellets were resuspended in 50ml of lysis buffer (Appendix 6). The tube containing the cell suspension was put on ice and a sonication probe placed into the solution, which was sonicated at 20 amps with 2 second pulses for 1 min. The solution was then centrifuged at 5000rpm for 15 mins at 4°C. The supernatant (soluble fraction) was removed and kept in a fresh tube and the pellet (insoluble fraction) was resuspended in 5ml of lysis buffer. 15µl aliquots of these fractions were run on an SDS acrylamide gel (Section 6.2.5.2) to determine whether proteins were soluble (Fig 6.7a), as GdMYB8b and GdMYB8c were the most

soluble they were used for the next steps. The supernatant was incubated with an Ni-NTA resin pre-balanced in loading buffer (Appendix 6), loaded onto a column, and the flow through (FT) was collected and kept. Two wash steps were completed that involved running a buffer through the column and then keeping the flow through, the first wash step used loading buffer (W1) and the second used washing buffer (W2) (Appendix 6). Finally, elution buffer (25ml base buffer and 50µl 1M DTT) was added 1ml at a time so that each elution fraction had a 1ml volume. 15µl of each fraction (insoluble, soluble, FT, W1, W2 and each elution) was prepared and run on a SDS acrylamide protein gel (Section 6.2.5.2) (Fig 6.7b). Tubes containing the elution fraction that contained the majority of the protein were pooled together to create one tube of GdMYB8b purified protein and a second tube of GdMYB8c protein. These samples were placed in individual dialysis membrane cassettes (Slide-A-Lyzer, 0.5 - 3ml capacity, Thermo Scientific) which were partially submerged in a dialysis buffer (as per the manufacturer's instructions) and incubated at 4°C overnight. The contents of the cassettes were then aliquoted, snap frozen in liquid nitrogen, and stored at -80°C. These aliquots were used in gel shift assays.

6.2.7.4 Identifying potential binding motifs

Potential binding motifs in upstream regions of *GdANS*, *GdDFR*, and *GdMAT1* were determined by searching within sequences for the motifs identified by Kelemen et al. (2015) that bind *A. thaliana* subgroup 6 R2R3 MYB transcription factors in yeast one-hybrid experiments. These motifs were cross-referenced with that of the binding motif found for *AtMYB113* (subgroup 6 MYB) in an *A. thaliana* cistrome (O'Malley et al. 2016). All *A. thaliana* motifs from this literature contained the sequence GTTA/G. As such, the criteria decided for potential GdMYB8 binding motifs were sequences that had no more than 2 SNPs different from *A. thaliana* motifs and that contained GTTA/G. A list of predicted motifs in the upstream regions of genes encoding *G. diffusa* anthocyanin synthesis enzymes was created. Motifs were also selected from the Kelemen et al. (2015) paper to be used as positive and negative controls to determine whether GdMYB8 proteins could bind, indicating that protein purification had worked and proteins were intact. For each control and *G. diffusa* predicted motif a 19bp region was isolated with the motif in the centre and a G nucleotide was added to the 5' end of the sequence. This 20bp oligo sequence was ordered along with a 19bp (excluding the extra G) reverse complement from IDT (idtdna.com).

6.2.7.5 Gel shift assay

Pairs of oligo sequences were annealed by mixing 20µl oligo 1 (40µM), 20µl oligo 2 (40µM), 5µl 10x annealing buffer (Appendix 6), 5µl ddH_2O, and heating at 96°C for 6 mins with a gradual cooling down to room temperature. The annealed oligos were then diluted to 0.8µM concentrations and 5µl were added to 2µl 10x Klenow buffer, 1µl Cy3-dCTP (8µM), 1µl Klenow, 11µl sterile ddH_2O. The solution was mixed by pipetting, incubated at 37°C for 2 hrs and then 65°C for 10 mins. For each set of annealed oligos two mixtures were prepared to be loaded onto the gel: one with the GdMYB8 protein added and one without protein. 1µl of annealed and labelled oligos was added to 2µl 10x fish sperm DNA (final concentration 10µg/ml), 13/17µl binding buffer (Appendix 6), 2µl DTT, and 0/4µl GdMYB8 protein. The solutions were carefully mixed through pipetting and left in ice for 30 mins to allow protein binding before being loaded onto an acrylamide gel.

An acrylamide protein gel was made (Appendix 6), differing from previously described gels in that it did not contain SDS (not a denaturing gel). 20µl of each sample were loaded and the gel was run in

0.5x TBE buffer for 30 - 90 mins at 90V. The gel was imaged on a Typhoon biomolecular imager with fluorescence acquisition mode and cy3 wavelength (550nm).

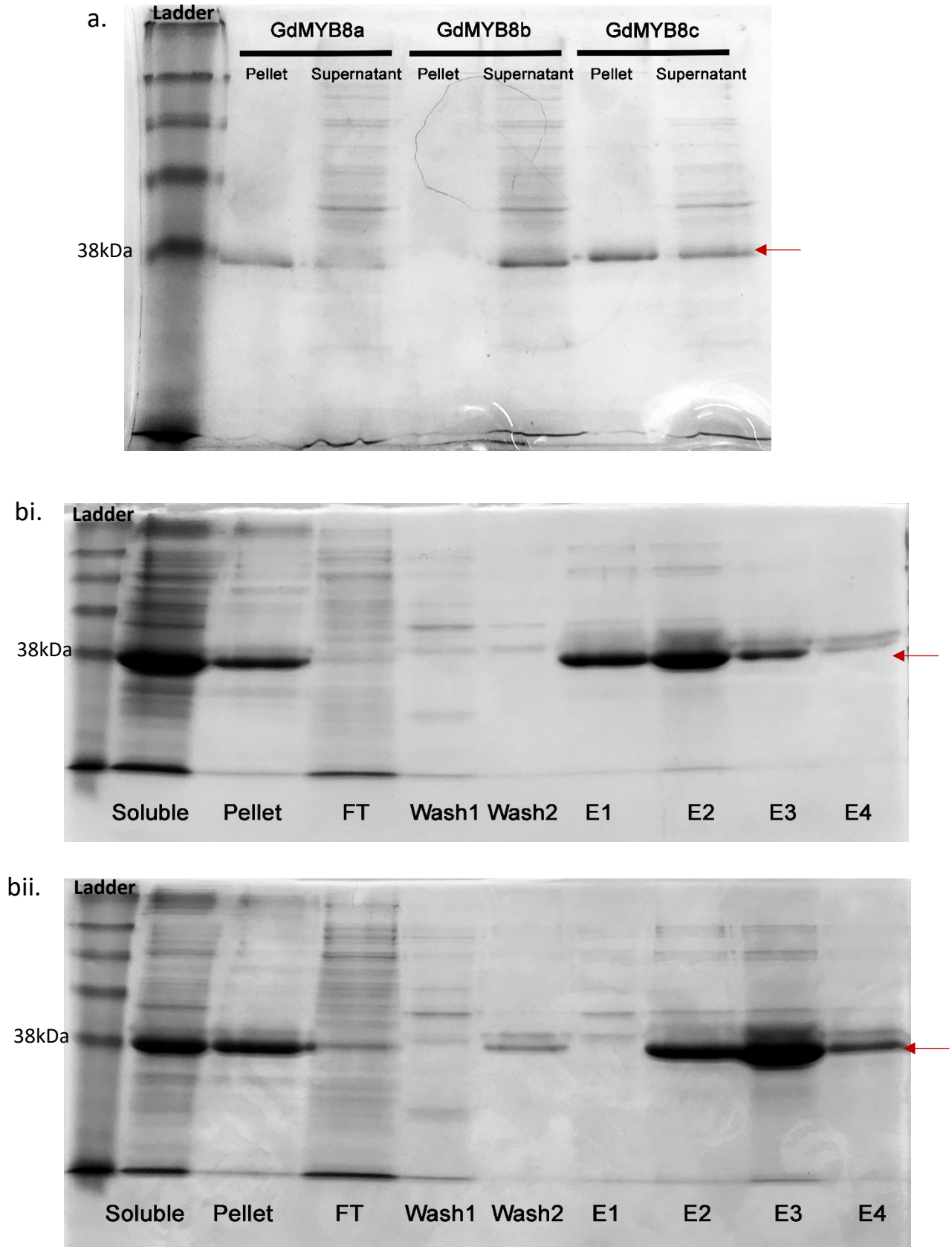


Figure 6.7 SDS acrylamide protein gels to determine whether a) GdMYB8 proteins were soluble, b) which wash/elution fraction contained the majority of the protein i) GdMYB8b ii) GdMYB8c. 'Soluble' is the supernatant prior to purification, 'Pellet' is the pellet prior to purification, 'FT' is the flow through of the column, 'Wash 1' and 'Wash 2' are the wash steps, 'E1 - 4' are the elution steps.

6.3 Results

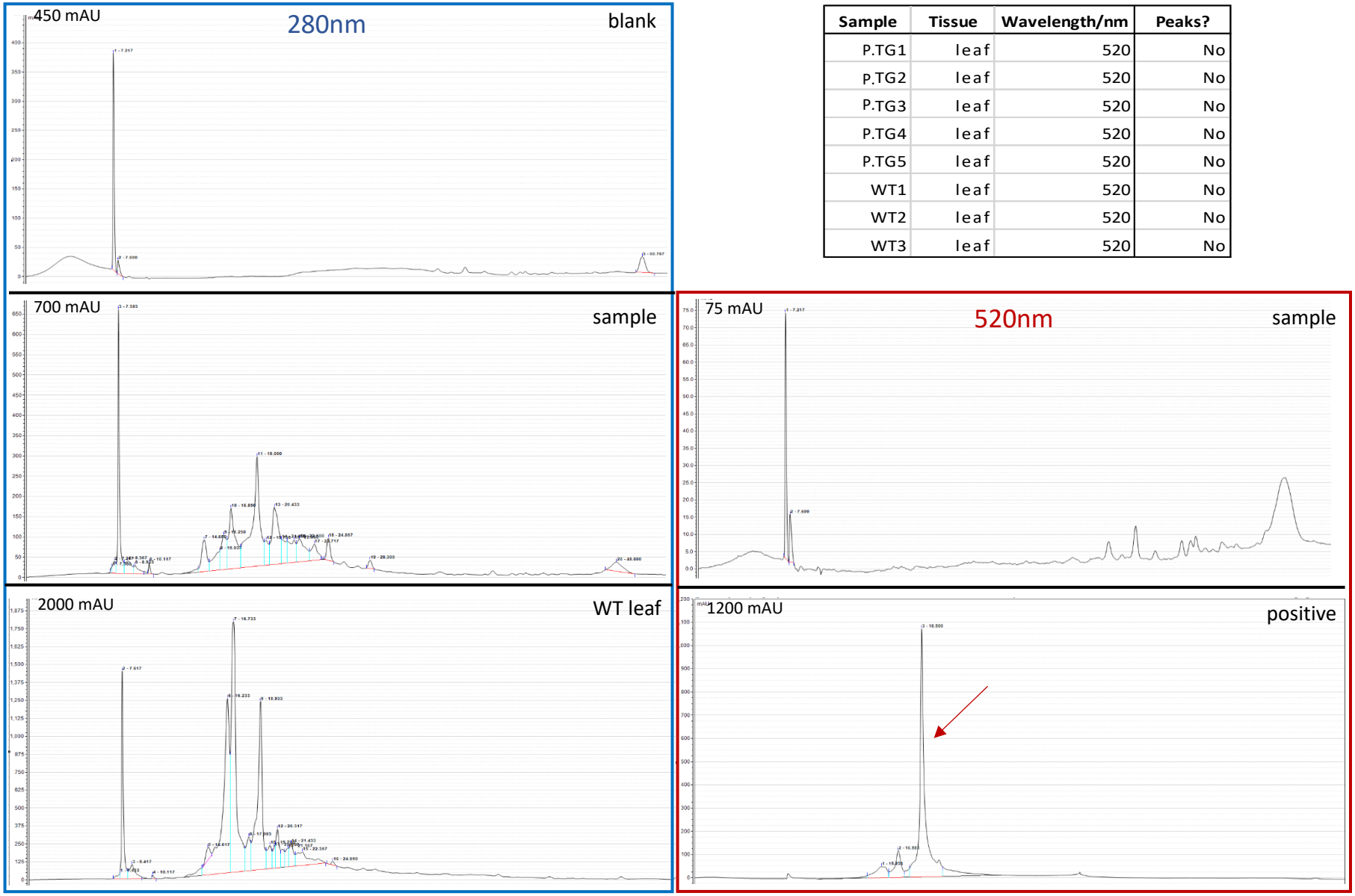
6.3.1 *G. diffusa* stable transformation trial

G. diffusa was stably transformed by Mellers (2016) with a vector containing 2x35S::Venus-NLS (a fluorescent protein) and hygromycin selection. Transgenic plants expressed Venus in mature floral tissue and leaves (Fig 6.8aii). The floral phenotype of the transformed plants was different from that of wild type capitula (Fig 6.8ai). To determine whether these phenotypic effects were due to stress imposed by regeneration or heritable changes, assessment of floral phenotypes of transformed plant progeny was required. Crosses between transformed plants, and between transformed plants and wild type Spring plants, were conducted over a 4 month period. Subsets of capitula were also netted following hand-pollination to exclude herbivores that may consume pollen. Unfortunately, these attempts were unsuccessful. Cuttings were taken to maintain the transformed plant lines and the vast majority of capitula on these plants (n = 5) contained no spots, with the exception of two capitula on one plant that had a single spot (Fig 6.8b).

Figure 6.8. G. diffusa plants stably transformed with Venus-NLS on a constitutive promoter by Mellers (2016). a) Images taken from Mellers (2016) i) The transformation process changed the spot and ray floret phenotypes compared to wild type ii) Petal spots expressed the nuclear localized protein. The top photo is taken in bright field and the bottom photo with GFP2 filter. b) Cuttings taken from transformed plants had no spots (n = 5), with the exception of 2 capitula on 1 plant that had 1 spot each.

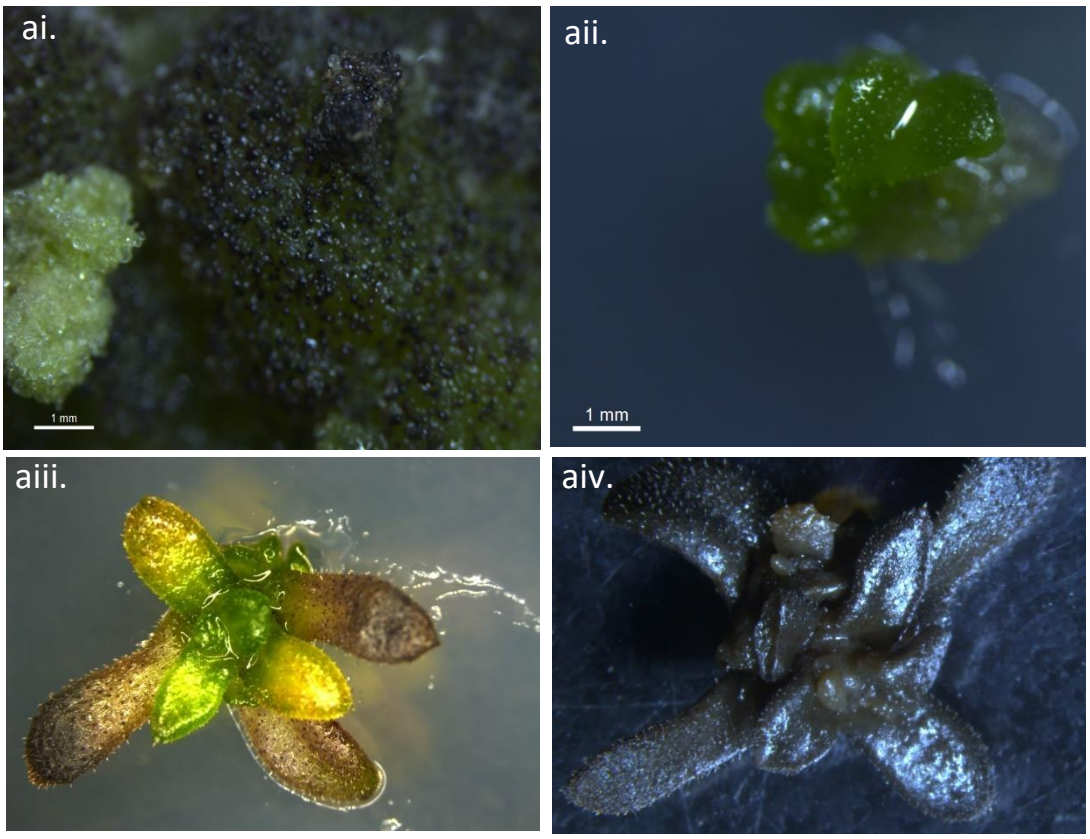
G. diffusa plants were transfected with a vector containing *GdMYB8a* (Fig 5.1, Section 5.2.3) under a constitutive 35S promoter. The first leaf emerged after 5.5 months, with leaves emerging from multiple calli after 8 months. Some plantlets had black pigmented leaves (Fig 6.10) - the phenotype expected for anthocyanin expression. Of the plantlets thought to be expressing anthocyanin, some were heavily pigmented since emergence, whereas others were initially green and black pigmentation occurred gradually from the leaf tip downwards as the plantlets matured (Fig 6.10). Individual plantlets were moved to Hamilton jars (n = 80) and those that were in media with no antibiotics and appeared green (no dark pigmentation visible) rooted. Once there was root development and sufficient growth of plantlets, they were transferred from Hamilton jars to soil, this took 12 - 13 months from the initial transformation (i.e. the infection of leaf discs with the vector of interest). These plants (n = 5) were expressing the transgene but had both floral and vegetative phenotypes that did not differ from wild type plants (Fig 6.10). As evident in Fig 6.11, wild type Spring plants have variable phenotypes and so the baseline for comparing transgenics in this experiment was unclear.

None of the plantlets with dark pigmentation rooted. Some plantlets were transferred to media containing auxin to see if this would stimulate root growth, but it did not and after several months no roots had formed. A subset of plantlets was selected (Fig 6.10b) representing a gradient of colouration (green to black). These samples were snap frozen in liquid nitrogen, ground, and split into two – with the intent that one sample would be used in HPLC analysis, to determine anthocyanin quantity, and the second in qRT-PCR to measure transgene expression level. This would enable testing for a correlation between pigment content and transgene expression level. However, acidic methanol extraction yielded a colourless liquid and HPLC analysis (Section 4.2.3) indicated that no anthocyanin was present in any of the plantlets (Fig. 6.9). Additionally, the heavily pigmented plantlets were dead (as expected) – confirmed by a lack of chlorophyll A and B peaks at the appropriate wavelengths in the HPLC analysis (data not shown). It was hypothesised that perhaps the pigment was not anthocyanin and instead *GdMYB8a* was inducing another type of pigmentation (e.g. tannins which are also derived from the flavonoid pathway). However, a *G. diffusa* transformation attempt by a postdoc in the lab using *GFP* (a fluorescent protein) to transform plants had similar colouration in the calli. It was concluded that the pigmentation seen here was probably a result of the regeneration procedure rather than transgene expression.



Sample	Tissue	Wavelength/nm	Peaks?
P.TG1	leaf	520	No
P.TG2	leaf	520	No
P.TG3	leaf	520	No
P.TG4	leaf	520	No
P.TG5	leaf	520	No
WT1	leaf	520	No
WT2	leaf	520	No
WT3	leaf	520	No

Figure 6.9. HPLC chromatograms of *G. diffusa* leaf material used to determine whether regenerated plantlets contained elevated levels of anthocyanin. In the blue box are chromatograms at 280nm (the absorbance spectra of flavonoids) for an acidic methanol blank, a transgenic plantlet (sample), and a wild type (WT) leaf. In the red box are chromatograms at 520nm (the absorbance spectra of anthocyanins) for a sample and a positive control (cyanidin-3-O-glucoside). Note that the anthocyanin peak (indicated by the red arrow) is absent in the transgenic plantlet. The y-axis scale differs between graphs so the maximum mAU (milli-Absorbance Units) is given as an indication of y axis scale. The table provides a summary of the HPLC analysis results. Five transgenic lines were analysed and 3 wild type leaf samples.



b.

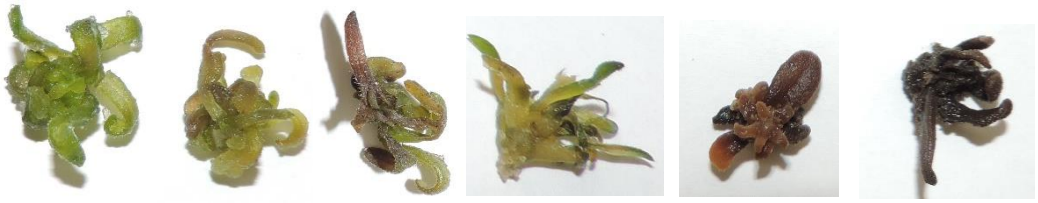


Figure 6.10. G. diffusa calli (ai) and plantlets (aii - aiv, b) from a stable transformation experiment. *GdMYB8a* was transformed into *G. diffusa* leaf discs on a constitutive promoter. Pigmentation was initially thought to be anthocyanin, (ai) first appearing in calli from which (aii) green plantlets formed. (aiii) Plantlets began to become pigmented or grew from pigmented sections of calli, but once this colouration spread plantlets died (aiv) with the colouration possibly instead due to necrosis. b) Example of samples used in HPLC to determine if anthocyanin could be detected.

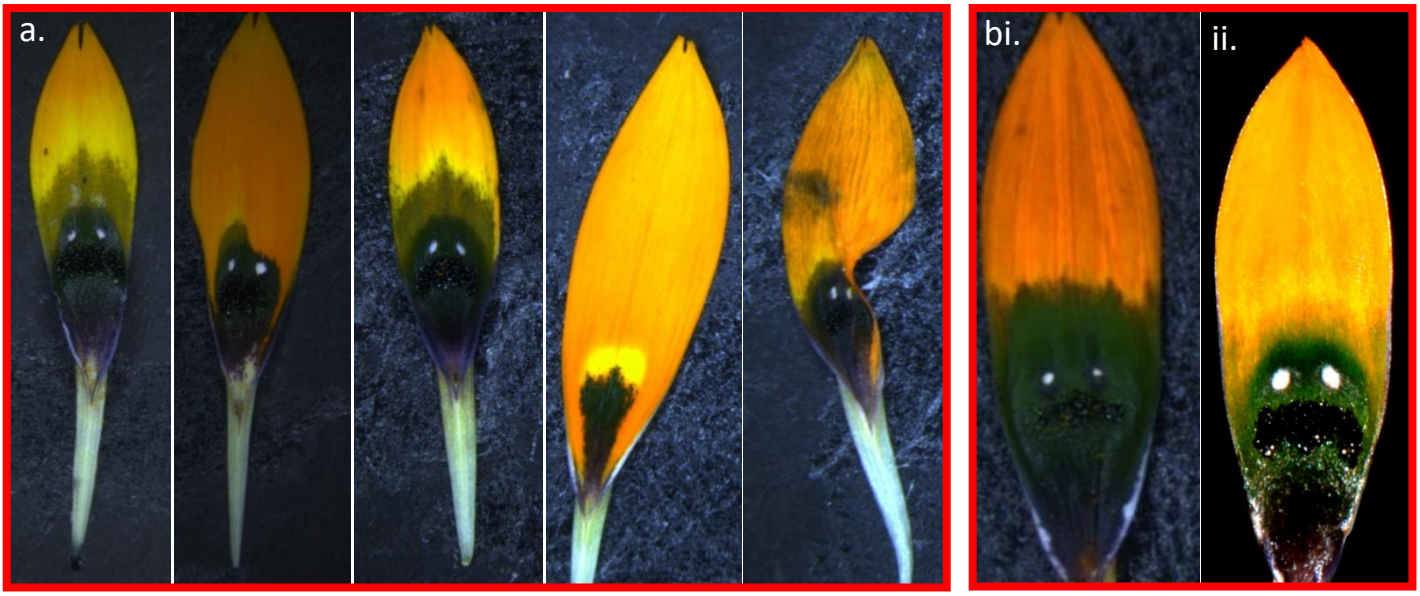


Figure 6.11. G. diffusa transgenic plants transformed with *GdMYB8a* on a constitutive promoter. a) Spotted ray florets from five T_0 transformants. Each plant had some completely developed spots and some partial spots – shown here are representatives of both. b) Wild type plants i) grown with the transformed plants ii) Spring ray floret presented as typical phenotype for the morphotype.

6.3.2 *G. diffusa* transient transformation trial

Attempts to develop transient transformation in *G. diffusa* were by no means exhaustive but did demonstrate that *G. diffusa* may be amenable to transient transformation with additional input. Petal tissue did appear to fluoresce due to infiltration (Fig 6.12b), while leaf tissue fluoresced only at the cut site (Fig 6.12a) with autofluorescence visible outside of the infiltrated site on leave hairs. Petal infiltration was difficult due to the delicate nature of the tissue and the majority of attempts damaged the petals causing them to curl (e.g. Fig 6.12bi). The damage to petal tissue often resulted in a darkening of the surrounding region, which could confound attempts to determine GdMYB8 functioning through transient transformation – as GdMYB8 is hypothesised to induce anthocyanin production. The size of fused ray florets meant it was not practical to trial transient transformation during different floral developmental stages. Leaves of different ages were used for infiltration, defined based on leaf size, and plants of different ages were infiltrated: during early vegetative growth, when plants were nearly full sized but not yet flowering, and flowering. These differences in parameter had no effect on the extent of transgene expression within infiltrated leaves. Due to the PhD timeframe additional trials were not conducted, but there are several other ways in which the protocol could be altered to try to develop a transient protocol for *G. diffusa*.

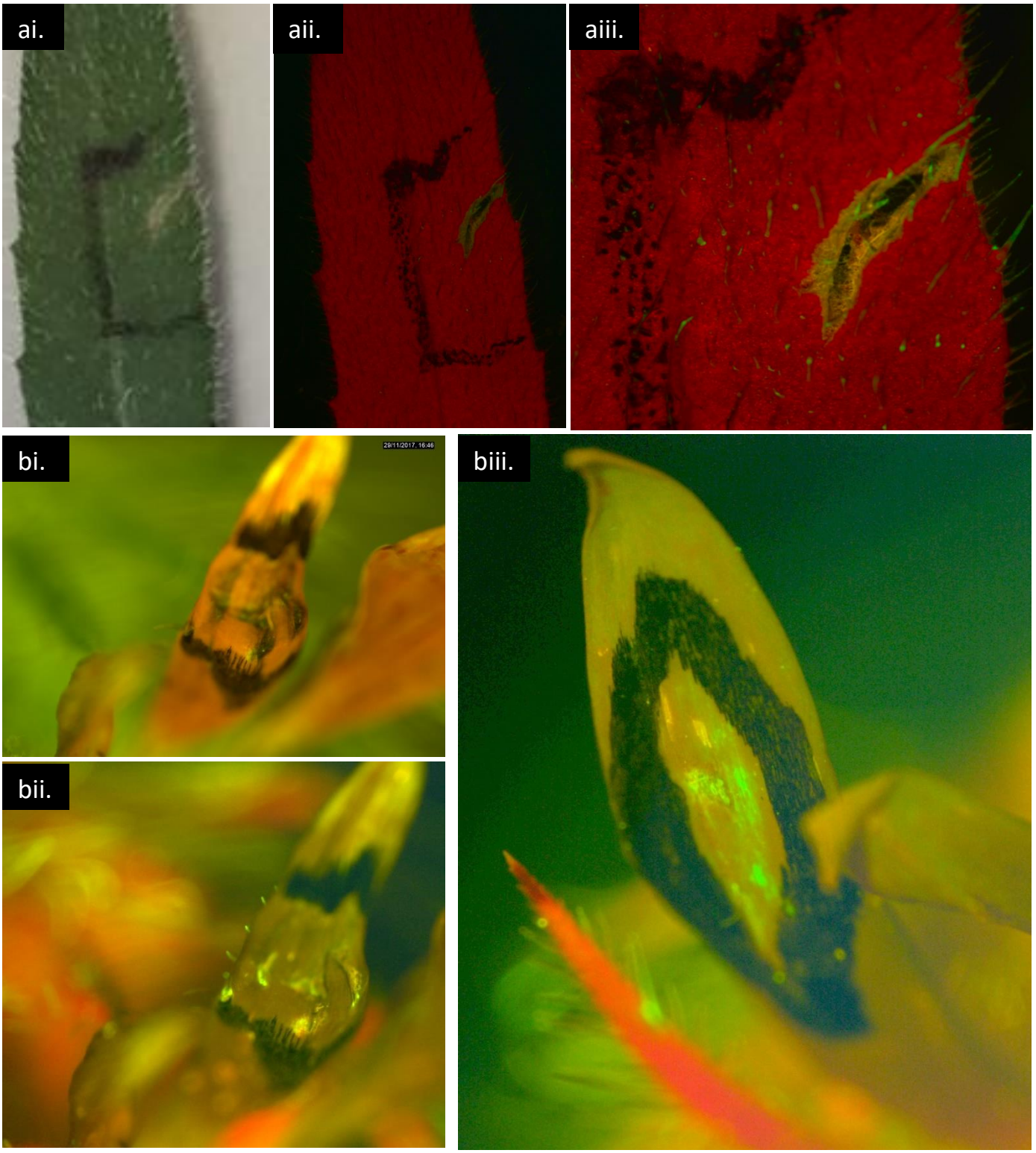


Figure 6.12. G. diffusa a) leaves and b) fused ray floret petals transiently transformed with *GFP* on a constitutive promoter. ai) and bi) were taken with no filter and a/bii) a/biii) were taken with a GFP2 filter. The black markings in each photo were drawn with a pen and represent the area within which the tissue was infiltrated.

6.3.3 Isolating promoter regions of *GdMYB8* genes

Upstream regions of the *GdMYB8* genes upregulated within spotted tissue were characterised using genome walking. The following lengths of promoter sequences were obtained: *GdMYB8a* 266bp, *GdMYB8b* 682bp, and *GdMYB8c* 673bp. These genes share similar expression patterns, potentially because they are regulated by the same transcription factor. If this is the case, then the binding site of this regulatory protein is likely to be within regions conserved between the promoters of each gene. Sequences were aligned to determine conserved sections (Fig 6.12). Within the first 266bp upstream of the start codon *GdMYB8a* was the most divergent compared to *GdMYB8b* and *GdMYB8c*, with an additional 27bp in one section and an absence of 30bp in another. Further upstream comparing *GdMYB8b* and *GdMYB8c*, *GdMYB8c* contains a section with an additional 70bp. Sections of *GdMYB8* upstream regions were also isolated in Cal and Stein (Appendix 6). These were Cal *GdMYB8a* and Cal and Stein *GdMYB8c*. All of these sequences were aligned and conserved regions were entered into the Plant Promoter Analysis Network (PlantPAN; http://PlantPAN.itps.ncku.edu/tw/)(Chow et al. 2019) to search for DNA binding motifs and predicted binding proteins, based on information from *A. thaliana*. The gene families of proteins that bind to identified *G. diffusa GdMYB8* motifs are listed in Table 6.2.

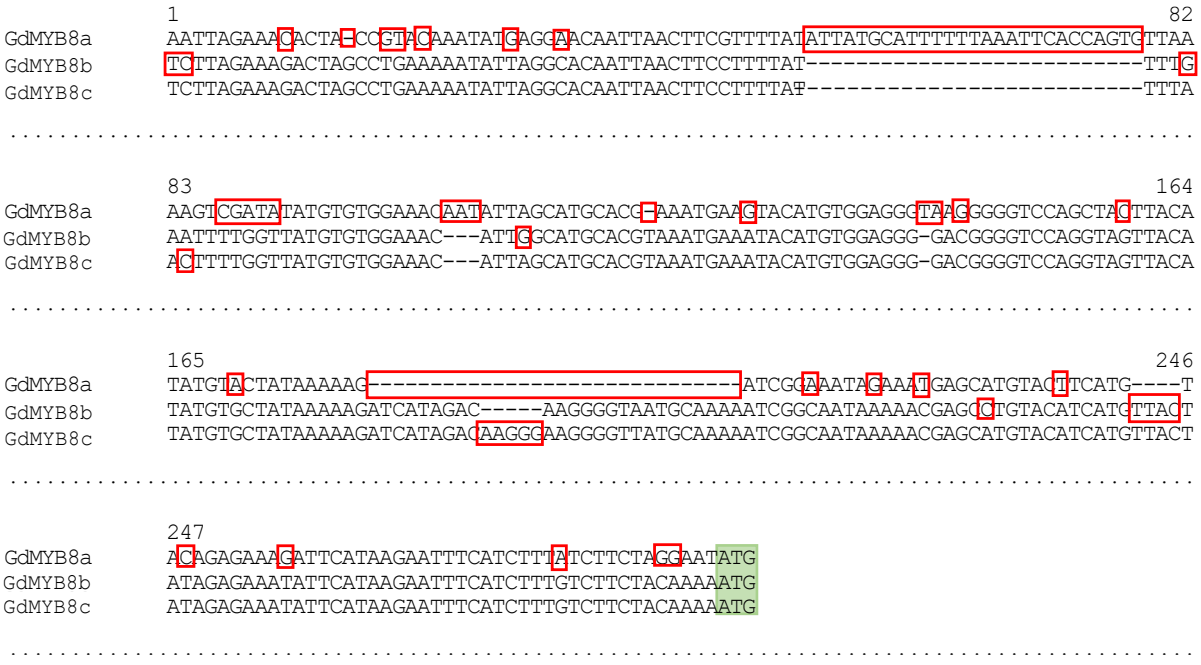
Protein family or type	Examples of regulatory functions	Reference
AP2 (APETALA 2)	Regulation of plant growth and development including floral organ identity, embryo development, and flowering time.	(Zhang et al. 2015; Licausi et al. 2013; Phukan et al. 2017)
ARID (AT-Rich Interaction Domain)	A superfamily involved in functions including nodule development, pollen tube growth, and shoot meristem development.	(Zhu et al. 2008; Xia et al. 2014; Xu et al. 2015)
B3 (Basic leucine zipper)	B3 transcription factors form a superfamily. They function in growth and developmental processes including control of embryogenesis, gynoecium development, floral morphogenesis, and abiotic stress responses.	(Braybrook and Harada 2008; Trigueros et al. 2009; Liu et al. 2014; Hu et al. 2015; Jain et al. 2009; Xia et al. 2019)
bHLH (Basic helix-loop-helix)	Regulate expression of many genes involved in multiple regulatory pathways including flower development, seed germination, root hair cell differentiation, light signalling regulation, and stress responses.	(Nadeau 2019; Oh et al. 2004; Menand et al. 2007; Kim et al. 2003; Sun et al. 2018)
bZIP	Seed maturation, light signalling, stress responses, and flower development.	(Abe et al. 2005; Strathmann et al. 2001; Alonso et al. 2009; Wang et al. 2018)
ZFP (Zinc finger proteins)	Flower development, stress responses, vegetative growth.	(Hu and Ma 2006; Huang et al. 2006; Mukhopadhyay et al. 2004; Colasanti et al. 1998; Liu et al. 2015b)
Dehydrin	Stress response – contribute toward plant protection against dehydration.	(Liu et al., 2017b)
Dof (DNA binding with one finger)	Regulation of defence responses, seed development, phytohormones, flowering, and germination.	(Dong et al. 2007; Rueda-Romero et al. 2012; Fornara et al. 2009; Nakano et al. 2006; Kang et al. 2003; Le Hir and Bellini 2013)
EIL; EIN3 (ethylene-insensitive3-like/ethylene-insensitive3)	Plant growth and development, and metabolic processes.	(Maruyama-Nakashita et al. 2006; Zhong et al. 2009; Guo et al. 2018; Binder et al. 2007; Salih et al. 2020)
GATA	Flowering time, flower development, embryo development, germination, lateral root initiation, and senescence.	(Behringer and Schwechheimer 2015)

PHD-Finger (plant homeodomain)	Type of ZFP. Seed germination, regulating flowering through chromatin modification, possibly pollen development and stress responses.	(Molitor et al. 2014; Fernández Gómez et al. 2014; Wei et al. 2009; López-González et al. 2014; Sun et al. 2017)
Homeodomain HD-Zip	Light and hormone signal transduction, stress responses, regulation of leaf polarity, apical meristem formation, vascular development, anthocyanin accumulation.	(Himmelbach et al. 2002; Gong et al. 2019; Sessa et al. 2018; Ariel et al. 2007; Wei et al. 2019)
MADS box	Central regulators of development, stress responses, integrators of endogenous hormones and environmental cues, primary metabolism, ABA signalling, detoxification processes, reactive oxygen species homeostasis.	(Castelán-Muñoz et al. 2019)
MYB	Secondary metabolism, cell cycle control, cellular morphogenesis, stress responses, hormone signalling.	(Gonzalez et al. 2008; Cominelli and Tonelli 2009; Ramsay and Glover 2005; Qi et al. 2011; Zhang et al. 2018)
NAC	Plant immunity, plant growth, and stress responses.	(Zhong et al. 2010; Bollhöner et al. 2012; Nakashima et al. 2012; Hussey et al. 2013; Yuan et al. 2019)
NF-Y (Nuclear-factor Y)	Embryo development, seed germination, photomorphogenesis, flowering time regulation, stress responses.	(Zhao et al. 2017)
SRS (Shi-related sequence)	Developmental processes including regulating signal transduction and hormone biosynthesis.	(Sohlberg et al. 2006; Eklund et al. 2010; Kim et al. 2010; Youssef et al. 2017; He et al. 2020)
TBP (TATA-box binding protein)	Required for every transcription event in eukaryotes and archaea.	(Sainsbury et al. 2015; Kornberg 2007; Rowlands et al. 1994)
TCP (Teosinte branched1/ Cycloidea/Proliferating cell factor)	Plant growth and development, cell growth and proliferation, regulation of seed germination, likely to play a role in plant immunity.	(Martín-Trillo et al. 2010; Schommer et al. 2008; Palatnik et al. 2003; Resentini et al. 2015; Bao et al. 2019)

Trihelix	Stress responses and developmental processes including sepal fusion, and leaf, petal, and sepal development.	(Fang et al. 2010; Xi et al. 2012; Xie et al. 2009; Brewer et al. 2004; Weng et al. 2012; Cheng et al. 2019)
WOX (WUS homeobox-containing)	Superfamily with family members that have specialised functions in plant developmental processes including organ development, embryonic patterning, and the maintenance of stem cells.	(Van der Graaff et al. 2009)

Table 6.2. The protein types for which binding motifs were found in the conserved upstream regions of *GdMYB8a*, *GdMYB8b*, and *GdMYB8c*. These findings are based on a search in PLANTPAN v3.0 against known *A. thaliana* binding sites. Examples of protein family functions are listed along with references. This provides a brief overview of potential candidate regulators. More precise exploration would require yeast-one hybrid experiments to determine which *G. diffusa* proteins are able to bind to these upstream regions.

a.



b.

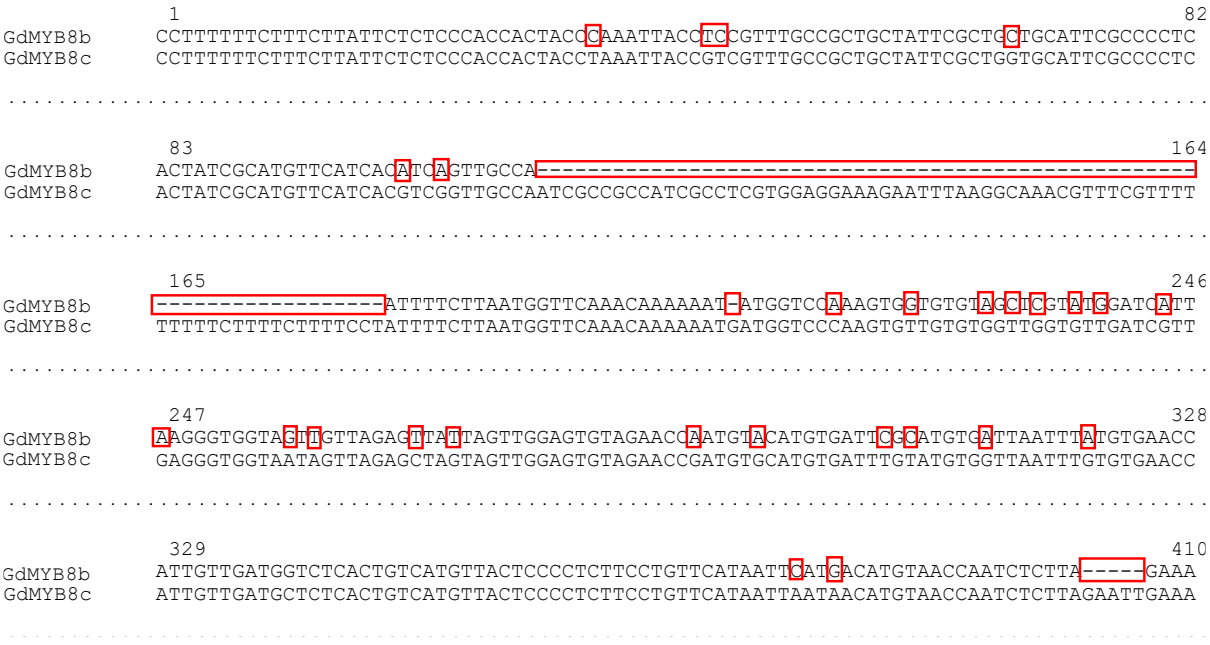


Figure 6.13. Alignments of promoter regions of a) *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* – the start codon is highlighted in green. b) Further upstream where only *GdMYB8b* and *GdMYB8c* sequences were available. Polymorphisms are indicated by red boxes.

6.3.4 Isolating promoter regions of *GdANS*, *GdDFR*, and *GdMAT1*

Upstream regions of each gene were isolated through genome walking and/ or primers designed from raw reads of a *G. diffusa* genome. 1763bp upstream of *GdMAT1* were sequenced in 6 individuals and was highly conserved between them, with only 17 SNPs. However, approximately 100bp upstream there was a 13bp insertion in 5 out of 8 sequences. This insertion contained part of a predicted MYB subgroup 6 binding motif; in the sequence lacking the insertion there is a different predicted motif that flanks the insertion region.

325bp upstream of *GdANS* was sequenced through genome walking - bands containing more of the upstream region were obtained but these could not be sequenced, presumably due to the repetitive nature of the sequence. The upstream region of *GdANS* was found in a raw read from an unassembled *G. diffusa* genome. This genome read was predicted to have an error rate of approximately 15% but was used for primer design to try to amplify the upstream region with a proof-reading enzyme. PCRs resulted in multiple bands per reaction with sequencing results that BLASTed to *ANS* and aligned with both the *ANS* gene and characterised 325bp upstream region. However, sequencing reactions failed, giving sequencing results that were too messy to be interpreted.

1945bp upstream of *GdDFR* was successfully sequenced and, due to the presence of multiple *GdDFR* variants, this upstream region was amplified several times across 8 individuals. The first 500bp upstream from the *GdDFR* start codon was found to be highly conserved – with just 14 SNPs differing and two short insertions (neither of which contained or disrupted a predicted subgroup 6 MYB binding motif). Further upstream was much more variable, for example, within the next 50bp upstream there were 15 SNPs, two 6bp insertions, and a 13bp insertion. Patches of conserved sequences were adjacent to highly variable regions. While it is unclear whether the upstream regions of all *GdDFR* variants were isolated, for all identified and loosely defined variants found here upstream regions were isolated.

6.3.5 GdMYB8 proteins interact with motifs in promoter regions of *GdDFR* and *GdMAT1*

Yeast one-hybrid experiments tested whether GdMYB8a, GdMYB8b, and GdMYB8c could bind to the promoter regions of *GdANS*, *GdDFR*, and *GdMAT1*. Promoter regions were divided into fragments of approximately 200bp (Fig 6.14) and used to generate the bait strains. In total, 6 bait strains of *GdMAT1*, 2 bait strains of *GdANS*, and 4 bait strains of *GdDFR* were generated. The conserved promoter regions of *GdDFR* were within fragments 1 and 2, while the remaining fragments contained promoter regions specific to a subset of *GdDFR* variants. Predicted binding motifs within bait sequences were identified (Fig 6.14, Fig 6.15, Table 6.3) as outlined in Section 6.2.5.4. This is an ongoing experiment, Y1H experiments have been completed for *GdANS* fragment 1, *GdMAT1* fragment 1-4 and 6, and *GdDFR* fragments 1 and 2 (Fig 6.15, Table 6.4). The remainder of the experimental work has now been taken over by Eva Herrero Serrano and Farahnoz Khojayori.

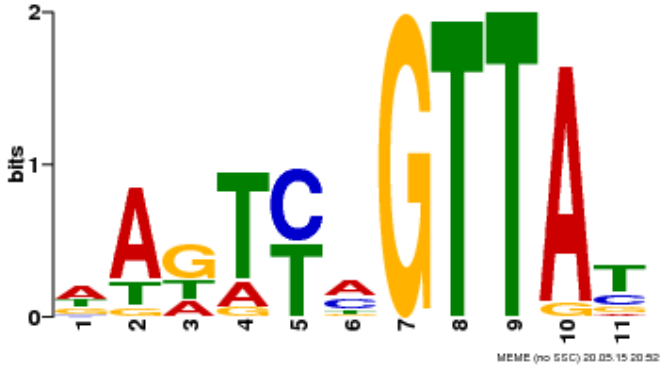
Several bait strains containing *GdDFR* and *GdMAT1* upstream fragments and prey-GdMYB8 plasmids were able to grow in 3-AT concentrations where prey-RVE1 and -VENUS containing strains did not (Fig 6.16, Fig 6.17, Table 6.4), demonstrating that these proteins can bind to motifs in *GdDFR* and *GdMAT1* promoters and activate *HIS3* expression. A positive result was indicated by colonies expressing GdMYB8 proteins that grew at greater 3-AT concentrations than the equivalent *Venus*-transformed control colonies and if this pattern was consistent across two independent transformation replicates

(Fig 6.16). All GdMYB8 proteins showed consistent binding patterns, although binding strength to some fragments differed between proteins. AtRVE1 bound to *GdDFR*-2 and *GdMAT1*-2, the former contained a RVE1 binding motif but the latter did not. There was one additional bait strain for which AtRVE1 binding was predicted (*GdDFR-1*), but no binding occurred. The only fragments for which no GdMYB8 protein binding was detected were *GdANS*-1 and *GdMAT1*-4, although for the latter weak binding was observed in one out of two replicates. The optimal 3-AT concentrations to eradicate background expression of *HIS3* varied between different bait strains. In some cases, the 3-AT concentration at which transformed colonies could grow varied between transformation replicates (possibly due to differences in 3-AT batches used between replicates). However, the relative concentrations at which the GdMYB8 transformed yeast could grow in comparison to controls generally remained the same. To confirm whether GdMYB8 proteins bind to predicted motifs, a subset of mutated fragments (representative of upstream regions from each genes) are being tested. Within these mutated sequences the conserved GTTA/G in each motif is replaced with GCCT/C (Table 6.3, Fig 6.14). If GdMYB8 proteins do not bind to these mutated upstream fragments, we can infer that one or more of these mutated sequences are the GdMYB8 binding motifs.

Regulators	Predicted motif	Source
Subgroup 6 R2R3 MYB	TACTGTTG	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	TGCGGTTG	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	AAAAGTTA	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	GTCAGTTA	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	ACAAGTTA	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	ATTAGTTG	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	GCTTGTTG	Kelemen et al., 2015
Subgroup 6 R2R3 MYB	AGTGGTTA	Kelemen et al., 2015
AtMYB113 (Subgroup 6)	NDDDYNGTTRN	O'Malley et al., 2016
AtRVE1 (MYB)	WNVNWAHKATCNNN	O'Malley et al., 2016

Table 6.3. Predicted subgroup 6 R2R3 MYB binding sites derived from *A. thaliana* research (Kelemen et al. 2015; O'Malley et al. 2016). The sequence GTTA/G is common to all of the motifs. RVE1 (REVEILLE 1) is a SHAQKYF-type Myb-like transcription factor that contains a single Myb-like domain (Rawat et al. 2009).

a.



b.

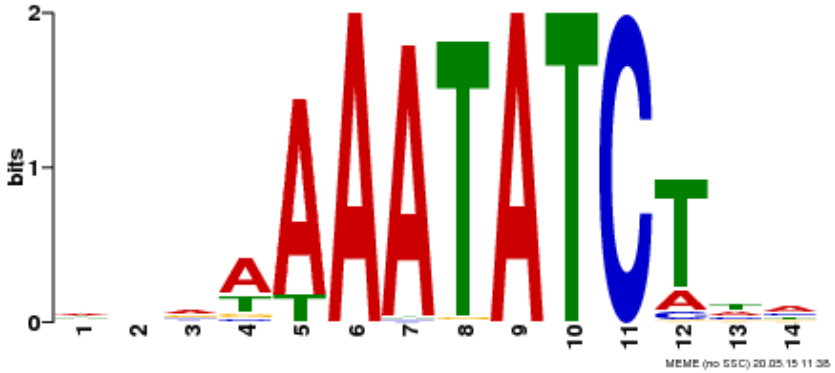


Figure 6.14. Binding motifs from O'Malley et al. (2016) for a) AtMYB113, an *A. thaliana* subgroup 6 R2R3 MYB and b) AtRVE1, a SHAQKYF-type Myb-like transcription factor.

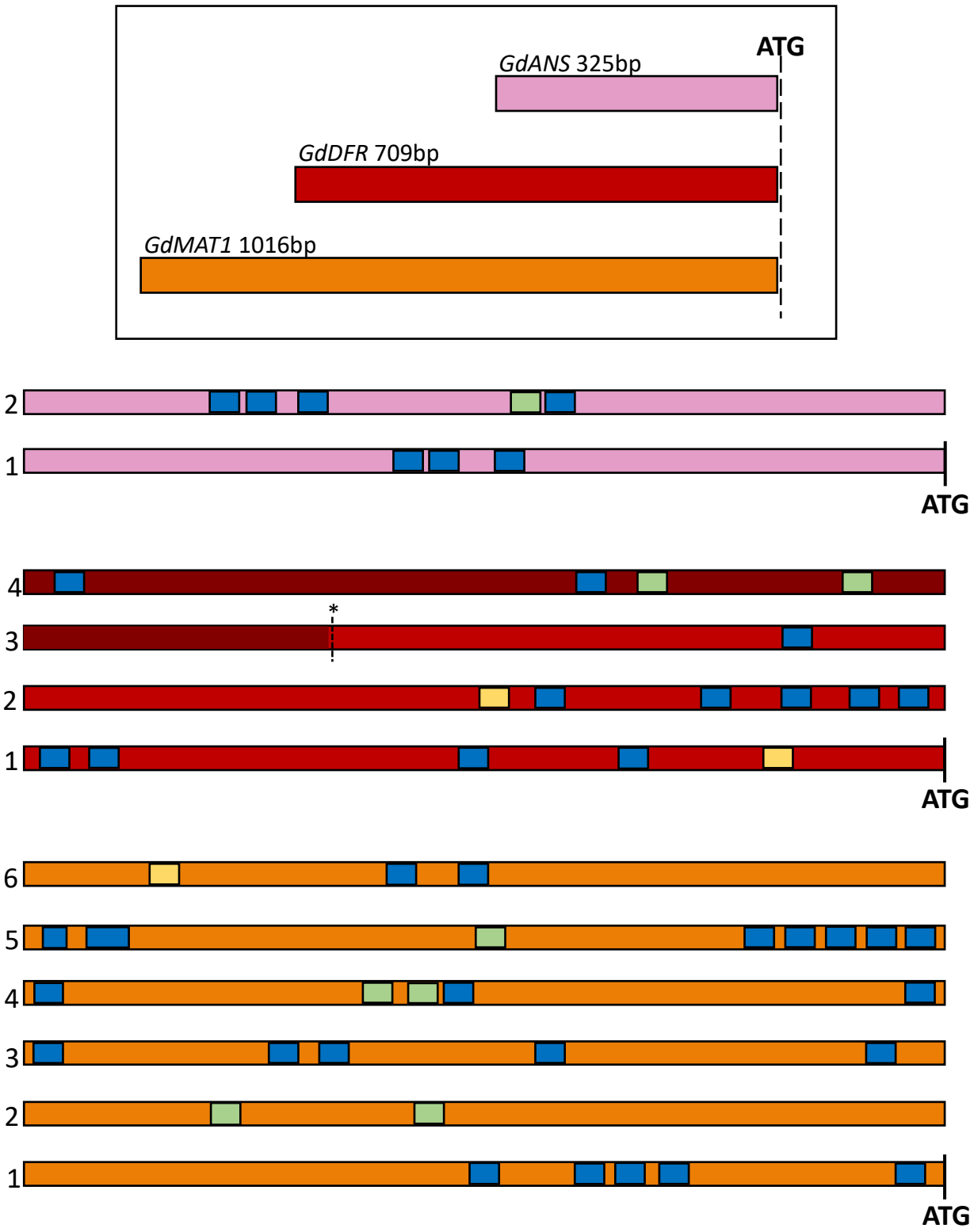


Figure 6.15. Schematic of the *GdANS*, *GdDFR*, and *GdMAT1* promoters. The box at the top illustrates the length of upstream regions that were used in yeast one-hybrid experiments. Promoter regions were divided into overlapping regions (fragments labelled 1 - 6) of approx. 150 - 200bp and the ability of GdMYB8 proteins to bind to each fragment was individually assessed. The 'ATG' is the start codon of each gene. Fragments are colour coded according to gene as illustrated in the top box. Predicted motifs identified (no more than 2 SNPs different from the reference motif and containing GTTA/G) are coloured in blue, motifs that the control MYB (AtRVE1) is predicted to bind to are coloured in yellow, and green motifs do not contain GTTA/G but only differ in 1 SNP from the reference motif. * indicates the point above which (further from the start codon) the *GdDFR* sequences become highly divergent between variants, indicated with a darker shade of red than the conserved *GdDFR* upstream region.

Promoter	Fragment	MYB8 binding	RVE1 binding	RVE1 motif present
GdANS	1	N	N	N
GdANS	2	n/a	n/a	N
GdDFR	1	Y	N	Y
GdDFR	2	Y	Y	Y
GdDFR	3	n/a	n/a	N
GdDFR	4	n/a	n/a	N
GdMAT1	1	Y	N	N
GdMAT1	2	Y	Y	N
GdMAT1	3	Y	N	N
GdMAT1	4	N	N	N
GdMAT1	5	n/a	n/a	N
GdMAT1	6	Y	N	Y

Table 6.4. A summary of the results of yeast one-hybrid experiments testing binding of GdMYB8 proteins to the promoter regions of anthocyanin synthesis enzymes (*GdANS*, *GdDFR*, and *GdMAT1*). All upstream fragments contained predicted subgroup 6 R2R3 MYB binding motifs and a subset contained RVE1 motifs presented in this table. n/a indicates that results are not yet available.

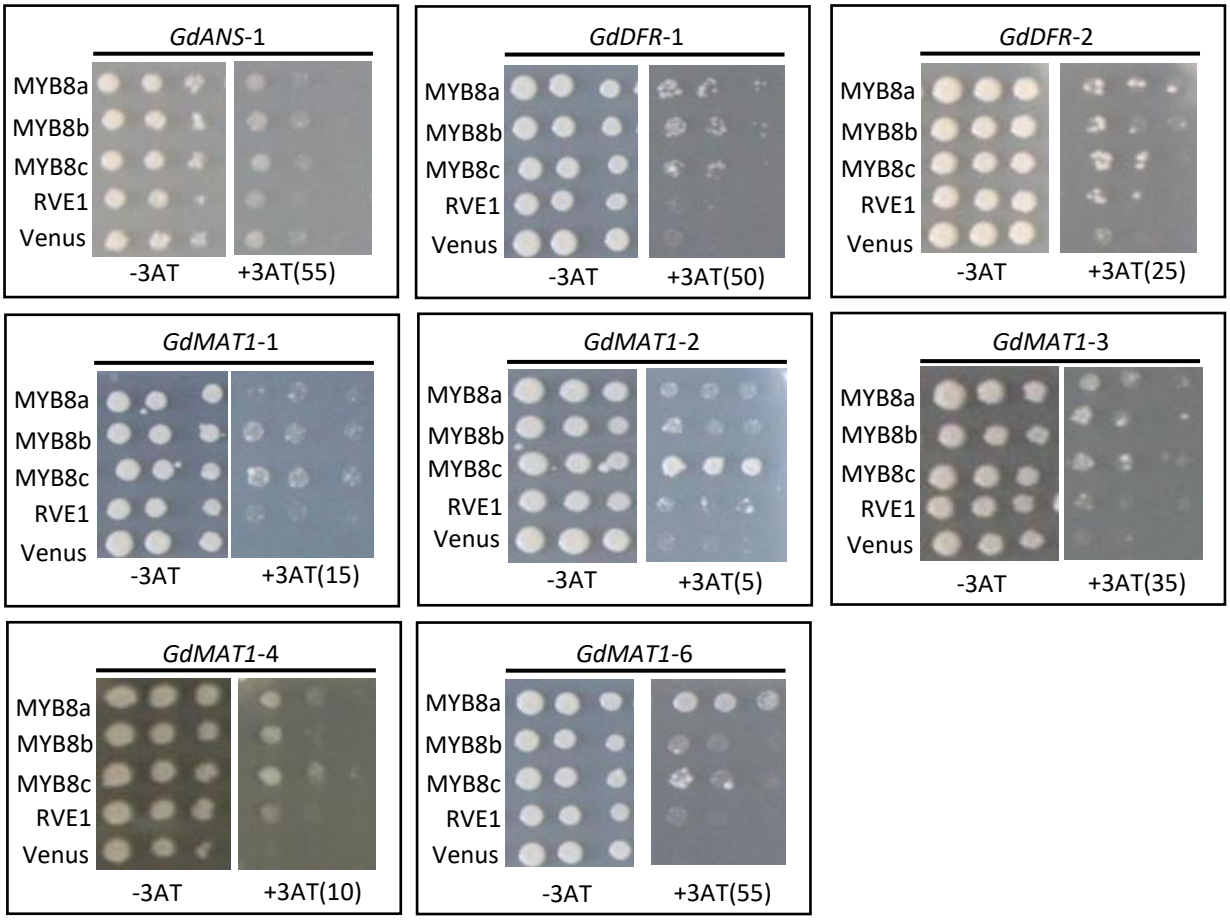


Figure 6.16. Photographs of yeast one-hybrid experimental results. The anthocyanin synthesis enzyme and upstream fragment number are indicated, corresponding to those illustrated in Fig 6.15. Each row corresponds to 3 dilutions of the same yeast colony transformed with either one of the *GdMY8* genes, *RVE1*, or the negative control *Venus*. -3AT plates lack 3-AT and demonstrate that all colonies are capable of growing on -HIS -LEU media. +3AT plates contain 3-AT at the concentration indicated in brackets (mM). GdMYB8 proteins bind to all fragment motifs with the exception of *GdANS*-1 and one replicate of *GdMAT1*-4 (not shown here).

***GdANS*-1**	Growth in Media lacking HIS LEU with 3-AT (mM)				
Frag 1	0	1	3	5	10
MYB8a					
MYB8b					
MYB8c					
RVE1					
Venus					

***GdDFR*-1**	Growth in Media lacking HIS LEU with 3-AT (mM)					
Frag 1	0	40	45	50	55	60
MYB8a						
MYB8b						
MYB8c						
RVE1						
Venus						

***GdDFR*-2**	Growth in Media lacking HIS LEU with 3-AT (mM)						
Frag 2	0	10	15	20	25	30	35
MYB8a							
MYB8b							
MYB8c							
RVE1							
Venus							

***GdMAT*-1**	Growth in Media lacking HIS LEU with 3-AT (mM)
Frag 1	0 15 20 25 30 35
MYB8a	
MYB8b	
MYB8c	
RVE1	
Venus	

***GdMAT1*-2**	Growth in Media lacking HIS LEU with 3-AT (mM)
Frag 2	0 1 3 4 5 10
MYB8a	
MYB8b	
MYB8c	
RVE1	
Venus	

***GdMAT1*-3**	Growth in Media lacking HIS LEU with 3-AT (mM)
Frag 3	0 10 15 20 25 30 35 40
MYB8a	
MYB8b	
MYB8c	
RVE1	
Venus	

***GdMAT1*-4**	Growth in Media lacking HIS LEU with 3-AT (mM)					
Frag 4	0	1	3	5	10	15
MYB8a						
MYB8b						
MYB8c						
RVE1						
Venus						

***GdMAT1*-6**	Growth in Media lacking HIS LEU with 3-AT (mM)					
Frag 6	0	40	45	50	55	60
MYB8a						
MYB8b						
MYB8c						
RVE1						
Venus						

Figure 6.17. Summary of yeast one-hybrid results demonstrating the 3-AT concentration at which each set of transformed yeast colonies were able to grow. GdMYB8 proteins or controls (Venus and RVE1) are listed, with growth at higher 3-AT concentrations than the Venus control indicating an interaction between the anthocyanin synthesis enzyme gene promoter fragment and the GdMYB8 protein. Yeast colony growth is indicated by blue bars, with each independent experimental replicate a different shade of blue. Underlined 3-AT concentrations indicate concentrations tested in one replicate only.

Gel shift assays were conducted to determine whether GdMYB8 proteins could bind to the predicted motifs in the promoter regions of *GdANS*, *GdDFR*, and *GdMAT1.* Using this method in combination with yeast one-hybrid results makes conclusions more robust and enables investigation into the specific motifs involved in protein binding. Unfortunately, purified GdMYB8 proteins were frozen at -80°C for a prolonged period of time during lockdown. Gel shift assays were conducted after lockdown and no binding was observed either to *G. diffusa* motifs or positive control motifs taken from Kelemen et al. (2015) (e.g. Fig 6.18). The proteins were run on an acrylamide gel to check for degradation induced by freezing, they did not appear degraded. Troubleshooting was conducted using several

different buffers obtained from a literature search (Appendix 6) and proteins were also left to bind to oligos at room temperature rather than in the cold. These methods did not work. Ultimately, it was concluded that it was likely the proteins were damaged from the freezing and thawing process and so fresh proteins needed to be produced and purified. This is a 3 day procedure and, as SLCU building access was limited to one day a week (due to COVID), it could not be completed. As such, the experiment remains unfinished.

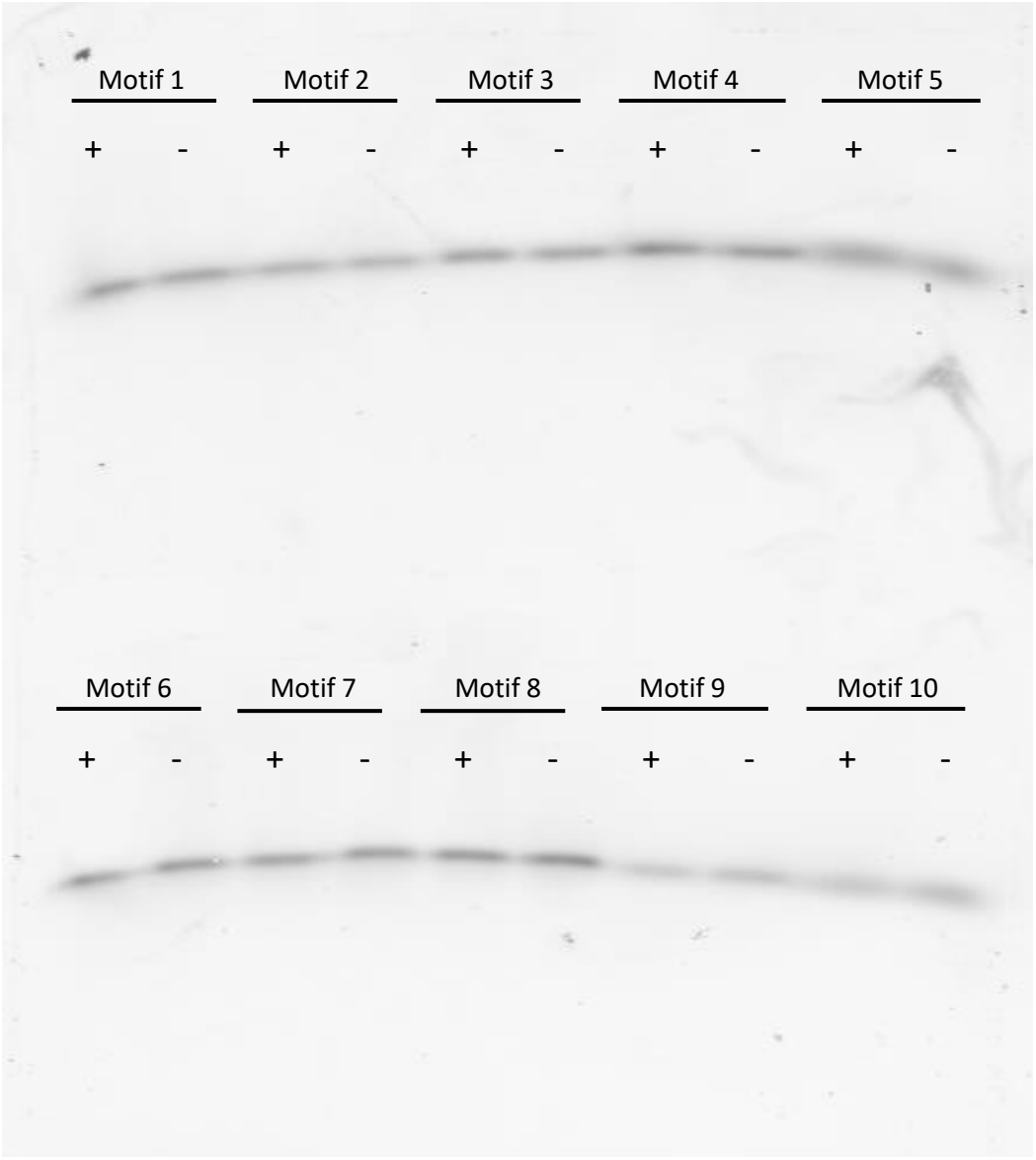


Figure 6.18. Gel shift assay on an acrylamide gel. Bands are fluorescently labelled oligo sequences containing GdMYB8 predicted binding motifs from protomer regions of genes encoding anthocyanin synthesis enzymes. '+' indicates GdMYB8b protein has been added and '–' is the negative control with no protein added. No binding was detected.

6.4 Discussion

Several approaches were used to try and determine the function of GdMYB8 proteins in *G. diffusa*. Stable transformation of *G. diffusa* with *GdMYB8a* did not yield conclusive results. While initial trials indicated that *G. diffusa* may be amenable to transient transformation, it was not possible to develop a transient transformation protocol within the project timeframe. Yeast one-hybrid (Y1H) experiments demonstrated that GdMYB8a, GdMYB8b, and GdMYB8c were capable of binding to promoter regions of *GdDFR* and *GdMAT1* to activate transcription of a reporter gene within yeast. Investigations into the ability of these proteins to bind to *GdANS* promoter regions are ongoing, as are validation of the specific motifs GdMYB8 proteins bind to. The latter is being conducted through Y1H experiments using mutated promoter fragments and gel shift assays to test for binding to individual predicted motifs.

The stable transformation of *G. diffusa* with *GdMYB8a* on a constitutive promoter yielded transgenic plants that were grown to maturity and expressed *GdMYB8a*, but the phenotypes of mature transformants did not differ from wild type. Five transgenic plants were grown from two independent lines. There are several potential explanations for the wild type appearance of transgenic plants including that GdMYB8a is non-functional in *G. diffusa*, expression levels of the transgene are too low to produce a phenotype, or post-transcriptional gene silencing within floral tissue prevented protein production. None of the regenerated plantlets with dark pigmentation rooted, and the accumulating pigment was not anthocyanin. *G. diffusa* calli transformed with GFP (by another lab member) had similar black/brown colouration suggesting that the pigmentation was not a result of transgene expression. Necrosis of plant tissues has occurred during stable transformation of other species using *Agrobacterium*, including maize and a gradual spreading of necrosis in grapes (Deng et al. 1995; Hansen 2000; Pu and Goodman 1992). Given that the most heavily pigmented plantlets were dead, it is likely that the pigmentation was a product of necrotic tissue. Laitinen et al. 2008 stably transformed *G. hybrida* using a 35S-*GMYB10* construct that resulted in accumulation of anthocyanin in floral tissues. However, they reported that most calli did not form shoots and postulated that this was due to toxic effects induced by excess anthocyanin. The pigment content of the calli shortly after calli formation should have been analysed to determine whether anthocyanin was accumulating in calli initially or whether dark pigmentation within calli was due to tissue necrosis, even during these early growth stages. If excess anthocyanin was having toxic effects in calli tissue it is possible that only plant cells with very weak transgene expression were able to differentiate into plantlets. This would provide a potential explanation as to why mature transgenic plants had wild type phenotypes, if only those weakly expressing the transgene survived. To resolve this conjecture would require further transformation attempts.

Important insights were gained from trialling stable transformation in *G. diffusa*. Increasing concentrations of cytokinin relative to auxin in regeneration media, relative to previous attempts (Mellers 2016), did not improve regeneration time. To determine whether the lengthened regeneration time reported here was due to hormone concentrations would require multiple experimental replicates; ideally transformations would be repeated with several different hormone treatments conducted in parallel. This was attempted by transforming *G. diffusa* and then splitting infected leaf discs into two sets that were grown in parallel but regenerated on media containing different hormone concentrations. Unfortunately, these plates were overcome with *Agrobacterium* infections and all calli died after three months, before regeneration had occurred (data not shown). The phenotypic variation seen between wild type plants grown alongside transformants and other wild type plants, grown over the course of the project, prompted a phenotypic characterisation of the

Spring morphotype across its natural range (Chapter 3). Understanding the extent of variability in natural floral phenotypes, and whether patterns of variation occurred between populations, could be useful for comparisons between transgenic and wild type plants. These analyses would also provide information on whether the geographical locations from which seed stocks are collected are important for ensuring phenotypic consistency in developmental analyses. Attempts to cross transgenic plants (transformed with GFP by Mellers (2016)) with one another and with wild type plants failed. Cuttings taken from these transgenic plants had no spots (with the exception of one plant that had a single simple spot on two capitula). Root growth of cuttings was stimulated by dipping severed shoots in an auxin powder, so these plants were subjected to elevated hormone levels which may have resulted in the phenotypic alterations. However, using this procedure on wild type plants has no impact on the floral phenotype of mature plants grown from cuttings. Wild type Stein individuals can also simultaneously produce capitula containing spots and capitula completely lacking spots. The combination of transgenic floral phenotypes and variation in spot presence, between capitula of wild type plants, implies that there may be epigenetic factors influencing spot development. Cytosine methylation, for example, is a key plant epigenetic mechanism that influences gene expression, the activity of transposons, and plant development (Cokus et al. 2008; Finnegan et al. 2000; Lister et al. 2008). To test this hypothesis, levels of DNA methylation could be measured between organs within Stein plants to see if heterogeneity in DNA methylation patterns correlate with the presence or absence of spots on capitula (Alonso et al. 2018; Herrera et al. 2019, 2020).

Informed recommendations for future *G. diffusa* stable transformation attempts can be made based on these experiments. The phenotypes of transgenic plant progeny must be assessed. Given the low efficiency and long regeneration time required for *G. diffusa* stable transformation, it would be advisable to conduct large-scale crossing experiments with wild type plants. A useful modification was recently suggested by our collaborator, who thinks that the stigma may be more receptive to pollen at an earlier developmental stage than that used for our crossing attempts. *G. diffusa* seed germination is erratic, and some seeds can remain dormant for one to several seasons (Duncan and Ellis 2011). As such, developing a successful crossing protocol able to yield many seeds per plant may be necessary. In the wild, a single *G. diffusa* plant can produce dozens of infructescences containing multiple seeds (Ellis and Johnson 2010), so an efficient and optimised crossing system should circumvent issues relating to dormancy. Many ambiguities resulting from the current transformation trial could potentially be resolved if vectors used for transformation also included a marker transgene (e.g. *GFP*), enabling quicker identification of transgenic plantlets and more efficient screening of plants to determine if transgenes are expressed in tissues of interest. The problems of possible necrosis within calli and regenerated plantlets, coupled with alterations to plant phenotype potentially caused by the transformation process, indicate that robust controls are necessary. The following control lines could be regenerated alongside transgenic lines: plants infected by *Agrobacterium* containing an empty vector and plants not infected by *Agrobacterium* but regenerated from leaf discs. This would enable better assessment of whether observed phenotypes were due to transgene expression. There are also multiple ways in which regeneration time and transformation efficiency could be optimised; these include altering the types and concentrations of hormones used for regeneration, trialling different antibiotics, and changing growth conditions.

Upstream regions of *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* were isolated and provide a good future resource for identifying potential regulators of *GdMYB8* genes, once the *G. diffusa* transformation protocol is optimised. The *GdMYB8* genes upregulated within petal spots have similar expression

patterns between homologues and across morphotypes, and so may be controlled by the same regulator. As such, upstream regions conserved between homologues and morphotypes could contain binding sites for transcriptional regulators within the petal spot developmental pathway. Alternatively, these genes could be under the control of different regulators, which is why it is important to assess the function of potential regulators i*n vivo* through transformation of *G. diffusa*. The addition of *GdMYB8d* to comparisons between upstream regions could be very informative because this gene has expression patterns that differ from the other *GdMYB8* homologues. As such, regions common to all four *GdMYB8* genes could be excluded as they are unlikely to contain the *cis*-regulatory elements of interest. Currently, the *G. diffusa* genome is being assembled and this will provide an excellent resource for fast and efficient isolation of gene promoter regions for further comparative analyses. Using these promoter fragments in yeast one-hybrid experiments would enable identification of the *G. diffusa* proteins capable of binding to them. These proteins could then be investigated as candidate *GdMYB8* regulators.

Transient transformation of *G. diffusa* could be a good method to address whether GdMYB8 proteins regulate anthocyanin synthesis. Preliminary transient transformation trials with a fluorescent protein did yield patchy transgene expression within transfected areas. Damage to *G. diffusa* ray floret petals can induce anthocyanin production (personal observation), but this was not the case in all attempted petal infiltrations. To prevent issues relating to tissue damage caused by cutting into epidermal petal layers, vacuum infiltration could be trialled. This method has been successfully used in transient transformation of *Petunia* flowers (Long et al. 2009). Although preliminary results were promising a *G. diffusa* transient transformation protocol could not be developed within the project timeframe to test GdMYB8 function.

Y1H experiments and gel shift assays were used to assess interactions between promoter regions of genes encoding anthocyanin synthesis enzymes and GdMYB8 proteins. These experiments demonstrated that GdMYB8a, GdMYB8b, and GdMYB8c all have the capacity to bind to promoter regions of *GdDFR* and *GdMAT1*. The 3-AT concentrations on which negative controls were able to grow were quite high. This is possibly due to insertions of the promoter fragments into the yeast genome occurring in tandem, increasing the level of leaky expression (where recognition of the bait sequence by endogenous yeast transcription factors cause gene activation) of *HIS3*. Further corroboration of these results is underway along with investigations into whether GdMYB8 proteins can bind to *GdANS* promoter regions. These findings provide further evidence that *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* likely encode proteins that regulate anthocyanin synthesis enzymes within the petal spots of *G. diffusa*. The identification of these regulators provides an entry point for investigations into the spot developmental pathway from which upstream spot regulatory components can be identified. Ultimately, in combination with additional genetic resources and tools for *in vivo* functional analyses, the results of this research could be used to begin unravelling the genetic networks responsible for other components of *G. diffusa* spot development.

Chapter 7. General discussion

Overview

Gorteria diffusa is a unique study system inhabiting the Succulent Karoo biodiversity hotspot. It exhibits high levels of geographically defined floral variation across a narrow endemic range. Unusually elaborate petal spots are major components of this floral variation, and in some morphotypes these spots induce mate-seeking and pseudocopulatory responses in the primary pollinator of *G. diffusa*, *Megapalpus capensis*. From a broad perspective, this is an ideal system in which to explore the ecological, evolutionary, and molecular developmental processes that promote divergence. Central to our molecular understanding is the characterisation of genes underlying petal spot development, while a prerequisite for evolutionary insight is to determine what comprises a morphotype in terms of genetic structure and phenotype. Here, the sexually deceptive Spring morphotype was used to conduct comprehensive investigations into population genetic structure and the regulation of petal spot anthocyanins. Molecular work was further corroborated through comparative analyses in the morphotypes Cal and Stein. In Chapter 3 population genetic studies and phenotypic characterisation were used to determine the spatial scale of gene flow and floral trait variation across the native range of Spring. The anthocyanin composition of petal spots was determined in Chapter 4 and a small family of subgroup 6 R2R3 MYB transcription factor genes (*GdMYB8*) were identified as good candidates for petal spot anthocyanin regulation through expression analyses. In Chapter 5 genes encoding anthocyanin synthesis enzymes were characterised and identified as potential downstream targets for GdMYB8 regulation, due to complementary gene expression patterns in petal spots. GdMYB8 proteins were shown to be capable of activating anthocyanin production in a heterologous system. Finally, Chapter 6 attempted to assess GdMYB8 functioning within *G. diffusa*, with promoter-binding assays providing further evidence that GdMYB8 proteins are capable of regulating anthocyanin synthesis within *G. diffusa*.

G. diffusa has multiple petal anthocyanin regulators

The fused ray floret petals of *G. diffusa* are pigmented by anthocyanin within the petal spots and across the abaxial surface. Pigment production is first initiated within petal spot primordia and abaxial pigment production temporally overlaps with the later stages of spot development. A small group of homologous subgroup 6 R2R3 MYB genes that were expressed in floral tissue were characterised. Three of these genes (*GdMYB8a*, *GdMYB8b*, and *GdMYB8c*) are thought to regulate anthocyanin synthesis within petal spots. These genes were upregulated in spotted petal ray floret tissue during spot development, they could activate ectopic anthocyanin production when stably transformed into *N. tabacum,* and the corresponding GdMYB8 proteins were capable of binding to promoter motifs of genes encoding *G. diffusa* anthocyanin synthesis enzymes. The fourth gene, *GdMYB8d,* was the most divergent in nucleotide sequence and is hypothesised to regulate abaxial anthocyanin pigmentation, as the timing of its upregulation in plain petal tissue corresponds with abaxial pigment production. These findings are consistent with those in several other systems, where multiple R2R3-MYB proteins were found to regulate anthocyanin biosynthesis in a single species.

Petal patterning often results from transcription of these regulators in a spatially and temporally distinct manner that dictates the petal anthocyanin pigment distribution; examples include *Lilium* (Yamagishi et al. 2014), *Petunia* (Albert et al. 2011), *Mimulus* (Ding et al. 2018; Yuan et al. 2014), and *Antirrhinum majus* (Schwinn et al. 2006). The expression patterns of *GdMYB8a*, *GdMYB8b*, and *GdMYB8c* indicate that the majority of gene transcription may be spatially restricted to petal regions

where the spot will develop. Alternatively, the genes could be expressed over a larger area, with activity inhibited in other petal regions through, for example, post-transcriptional gene silencing. Further characterisation of *GdMYB8d* would enable additional interesting comparisons to other systems, for example, if *GdMYB8d* does control abaxial pigmentation it may function like the R2R3 MYB PURPLEHAZE in *Petunia,* which regulates light-induced anthocyanin accumulation mainly on outer surfaces of the flower buds (Albert et al. 2011). These results add to a body of literature demonstrating that subgroup 6 R2R3 MYB transcription factors regulate developmentally programmed colour patterning in floral organs.

GdMYB8 proteins regulate genes encoding anthocyanin synthesis enzymes

While all anthocyanins detected within ray floret petals were cyanidins, those containing a malonyl group were detected almost exclusively in spotted petal tissue and comprised approximately 60% of petal spot anthocyanins across Spring and Cal. Malonic acid acylation can increase anthocyanin stability (Figueiredo et al. 1999), and the detection of these malonated anthocyanins within marks (simple patches of pigment at the base of plain Spring ray floret petals) indicated that this anthocyanin was common to all spot types, and not only those with cellular elaborations. The malonyl transferase *GdMAT1* had spot specific expression patterns and *GdMYB8* genes could bind to promoter regions of *GdMAT1*, indicating that it is a good candidate for spot-specific synthesis of malonated anthocyanins. In contrast, *GdANS* exhibits temporal differences in expression between spotted and plain petal tissue, with upregulation in developing spots preceding upregulation elsewhere. Investigations into whether GdMYB8 proteins can bind to *GdANS* promoter motifs are ongoing. *GdDFR* was also upregulated in spotted petal tissue at an early developmental stage and, unlike *GdANS*, at the later developmental stage investigated *GdDFR* expression was significantly higher in spotted compared to plain petal tissue. However, there are several *GdDFR* genes and individual expression patterns could not be determined. It is unclear whether there are *GdDFR* genes with spot-specific expression, like *GdMAT1*, or whether a temporal shift in expression is seen between tissue types, similar to *GdANS*. GdMYB8 proteins were able to bind to promoter fragments conserved between all *GdDFR* variants identified, implying that all copies could potentially be activated by GdMYB8 proteins within the spot. This requires confirmation from expression analyses considering each *GdDFR* gene or allele individually. There are often multiple copies of anthocyanin synthesis enzymes expressed within flowers. In Asiatic hybrid lilies (*Lilium*), for example, two *CHS* genes are expressed in the tepals, filaments, and pistils, while another copy is responsible for anther anthocyanin accumulation (Nakatsuka et al., 2003; Lai et al. 2012). The assembly of a *G. diffusa* genome (in progress) will provide important conformation of the number of *GdDFR* genes. With this more in depth understanding, it will be interesting to compare the expression patterns of various *GdDFR* copies to determine whether each gene or allele has a different role in ray floret petal pigmentation.

Divergence between *GdMYB8* genes

The large size of the MYB transcription factor family in plants has been, in part, associated with the rapid expansion of R2R3 MYB genes (Dubos et al. 2010; Jiang and Rao 2020). Duplication events of MYB anthocyanin regulators in *Mimulus* led to new floral pigmentation patterns. Within the five species of the luteus group of *Mimulus*, petal lobe anthocyanin has evolved in parallel in two species (*M. cupreus* and *M. luteus. variegatus*). Duplication of MYB genes occurred within different loci of each species that had similar functions (Cooley et al. 2011). Recent duplication in anthocyanin regulators is evident in several systems, including the grape *VvMYBA* genes (Walker et al. 2007), *A.*

majus genes *Ros1* and *Ros2* (Schwinn et al. 2006), and *AtMYB90*, *AtMYB113* and *AtMYB114* in *A. thaliana* (Gonzalez et al. 2008; Stracke et al. 2001). The duplication events producing the *G. diffusa GdMYB8* clade appear to have occurred fairly recently, but prior to *G. diffusa* speciation. Greater taxonomic resolution would be required to deduce more about the gene evolutionary histories. Several lineages of R2R3 MYBs have been found to contain recently duplicated genes that show signatures of positive selection (Jia et al. 2003). Following gene duplication, functional divergence of homologues through either neo or subfunctionalisation (e.g. Haberer et al. 2004) is one of several possible evolutionary scenarios, which also include functional redundancy of one gene copy and pseudogenization (Zhang 2003). Evidently *GdMYB8d* has diverged from the rest of the *GdMYB8* clade in terms of expression pattern; this gene was not upregulated during initial petal spot development and may not be expressed within spotted tissue at all, pending further investigation. There was divergence in expression levels between *GdMYB8a, GdMYB8b*, and *GdMYB8c* within Spring petal spots, and these differences were largely consistent within Cal petal spots. Overall, *GdMYB8b* and *GdMYB8c* had higher expression levels, suggesting that they may be the dominant genes in regulating petal spot anthocyanin production. Expression analyses could be conducted with greater spatial resolution to determine whether specific *GdMYB8* genes are localised to certain spot regions. One gene may, for example, have expression localised to papillae primordia cells indicating that they may trigger the increase in anthocyanin production seen specifically within this specialised cell type (Thomas et al. 2009). Investigating finer scale differences would also enable a more comprehensive comparison with *GdMYB8* expression patterns in other morphotypes.

Potential functional differences between GdMYB8 proteins were detected by stably expressing them in *N. tabacum*. Constitutive expression of all three genes induced strong anthocyanin phenotypes across several tissues, but *NtANS* upregulation was significantly lower in *GdMYB8a* transformants compared to *GdMYB8b* and *GdMYB8c* transgenic plants. As such, the GdMYB8a protein may have different biochemical properties than GdMYB8b and GdMYB8c. While the amino acid sequences of these three GdMYB8 proteins were relatively similar (84 - 91% homology), there were seven amino acid differences within the R2R3 MYB domains. This region is involved in the creation of a scaffold that enables insertion of the third helix (also encoded by this domain) into the major groove of DNA, where it binds to nucleotide bases (König et al. 1996; Ogata et al. 1992, 1994). Interactions with bHLH partner proteins are also mediated by a motif within this domain (Zimmermann et al. 2004). These differences in amino acids could have functional implications. Further insight into possible differences between GdMYB8 proteins, relating to the regulation of anthocyanin synthesis, might be gained from ongoing gel shift assays. These experiments will confirm whether or not all GdMYB8 proteins have the capacity to bind to certain motifs in the promoters of anthocyanin synthesis enzymes and allow comparisons of binding affinity.

Phenotypic characterisation contributes toward developmental understanding

G. diffusa has extreme intraspecific variation between floral morphotypes, including fixed differences in petal spot composition. The capitulum phenotype is also complex regarding additional variation within morphotypes, for example, the presence or absence of simple basal spots and the number of petal spots present within a capitulum (detailed in Fig 1.7). As such, single phenotypic measurements of *G. diffusa* individuals imply a static uniformity that may not be fully representative of floral phenotype (Harder et al. 2019). Understanding these dynamic phenotypes, and the levels at which variation occurs, is an essential prerequisite for certain developmental inferences and genetic

hypotheses. A thorough assessment of the scales at which variation occurs was undertaken in Spring and in a subset of Cal (Fig 7.1).

The morphotype Spring has a particularly complex *G. diffusa* capitulum phenotype. Specialised cell types within the spot must develop in the correct positions relative to one another, coordinated over four fused ray floret petals. Consistency in spot phenotype may be important for inducing mating responses in male bee-fly pollinators, with higher levels of phenotypic integration occurring between spot traits of sexually deceptive morphotypes (Ellis et al. 2014). Only a subset of ray florets within a capitulum produce petal spots and these ray florets have a distinctive shape, size, and colouration compared to ray florets which do not produce spots (plain ray florets). The number of spotted ray florets per capitulum can vary within an individual plant. *G. diffusa* ray florets develop basipetally and petal spots develop on the oldest ray florets, which are the most internal in the capitulum. Thomas et al. (2009) proposed that a developmental signal regulating petal spot fate is activated when the first ray floret is initiated, and this signal then slowly diminishes as additional ray florets are initiated. Prior to its dissipation, the signal would enable spot production in multiple ray florets – with the number of spotted ray florets determined by signal strength. Here, the number of spotted ray florets was found to be positively correlated with the total number of ray florets in a capitulum, which is also variable within an individual. This implies potential coupling between the strength of signals initiating the development of all ray florets and those inducing spotted ray floret formation. 26% of the capitula investigated (n=330) contained a 'partial spot' (n=86), where a complex spot phenotype had started to form but development was prematurely arrested. In the vast majority of capitula only a single partial spot was present (0.03% of capitula had 2 partial spots), supporting the proposed model of ray floret development. Interestingly, in ray florets with partial spot phenotypes the extent of spot development differed between the fused petals of the ray floret. Petals with more developed spot components (i.e. containing some of the specialised cell types) had phenotypes resembling 'spotted ray florets', shorter in length and lighter in colour than plain ray floret counterparts. Those petals with less developed spots (i.e. generally only pigment present) had petal phenotypes resembling plain ray florets. This demonstrates that spot development and ray floret identity are coupled during petal development. A larger sample size would enable more detailed inferences regarding the patterns observed.

G. diffusa is comprised of floral morphotypes, each of which consists of geographically grouped populations that have relatively consistent floral phenotypes. Between morphotypes there are distinct differences in various floral traits, resulting in discrete capitulum phenotypes associated with each floral form. The variation in spotted ray floret number between capitula within plants can be considered a characteristic that contributes toward the Spring floral phenotype (Kulbaba et al. 2017), potentially enhancing fly mimicry by creating a pseudo-random pattern of spots across the floral display as a whole. Spot and ray floret traits are highly consistent between individuals across the Spring range, with only minimum quantitative variation detected. The exception to this is the presence or absence of marks (simple spots) at the base of plain ray florets which varies between individuals; 76% of the plants investigated had marks present on the ray florets. Stein appears to have greater inherent variation, with developmental plasticity that enables a single individual to produce spotted and plain capitula simultaneously. In contrast, every ray floret and capitulum within Cal is consistently, and without exception, spotted by a complex petal spot (Fig 7.1). No partial spots have been observed in Cal plants. These findings have important implications for the use of *G. diffusa* as a genetic study system.

The proteins regulating R2R3 MYB transcription factors involved in floral patterning are largely unknown (Yuan et al. 2014). The promoter regions of *GdMYB8* genes were isolated as a first step in determining the regulators of these genes. As petal spot development involves several cell types and traits, there may be a limited number of spot regulator genes that coordinate different spot developmental modules. Comparisons between the promoter regions of *GdMYB8* genes, that are upregulated in the petal spot, enabled identification of conserved sequences. These sequences may contain *cis*-regulatory elements that petal spot regulatory proteins bind to. Yeast one-hybrid experiments would then enable identification of proteins capable of binding to motifs within these promoter fragments. The function of the candidate regulators could be assessed through stable transgenic analysis in *G. diffusa* by downregulating gene expression and inducing overexpression. A major current limitation is that stable transformation in Spring alters the spot phenotypes, and in cuttings of these transgenic plants no spots were produced (with the exception of 2 capitula in one plant). The plasticity of the spot makes phenotypic inferences from transgenic plants difficult, particularly those pertaining to whether or not certain genes can induce spot formation. Considering specific spot traits, the occurrence of partial spots within transgenic plants could also confound phenotypic comparisons. If wild type phenotypes are used as a proxy for suitability in transgenic experiments, Cal may be an ideal candidate because it exhibits none of the within individual and between individual qualitative variation seen in Spring and Stein (Fig 7.1). As such, the transformation and regeneration process may not alter the Cal phenotype. To test this hypothesis would require several rounds of transformation and regeneration with empty constructs, followed by phenotypic analyses of transgenic plants. If successful, this would enable downregulation of candidate petal spot regulatory genes within Cal, without the complications resulting from plasticity that is potentially a characteristic of some of the other *G. diffusa* morphotypes.

Using an interdisciplinary approach to understand diversity in *G. diffusa*

G. diffusa may be undergoing incipient speciation, evident from the geographically structured floral forms and recent work demonstrating genetic clustering of individual morphotypes (Boris Delahaie unpublished). The investigation into Spring population structure highlighted that isolation by distance is likely to contribute towards divergence. The high genetic differentiation present over small spatial scales indicates that limited dispersal may be an important mechanism enabling genetic differentiation between morphotypes over a narrow spatial range. Given that diversification of *G. diffusa* is evident through floral variation, it is likely that genes involved in flower development have undergone genetic divergence between morphotypes. In *Mimulus aurantiacus*, population genetics techniques were used to investigate patterns of genetic divergence between red and yellow ecotypes that exhibit a sharp geographic transition in floral form (Stankowski et al. 2016). Outlier scans were used to detect loci that were highly diverged between ecotypes, combined with a cline analysis that fitted allele frequency data to geographic cline models. Candidate loci (130) were identified that may contribute toward genetic divergence or reside in genomic regions near loci that do (Stankowski et al. 2016). Given that *G. diffusa* genome assemblies are underway, and genotyping by sequencing analyses are being conducted across morphotype hybrid zones, a similar approach may be attainable in *G. diffusa* in the near future.

Population genetic approaches could accelerate the identification of potential candidate genes involved in petal spot development, in parallel to experiments using *GdMYB8* genes as an entry point into molecular exploration of the petal spot developmental pathway. Ultimately, *G. diffusa* has much potential as a study system for investigating the genetics underlying complex floral patterning and the

evolution of floral divergence. The use of natural variation within the system as a basis for comparative frameworks and complementary interdisciplinary approaches should greatly aid these objectives. The work presented within this thesis provides an important foundation of knowledge, particularly within the Spring morphotype, from which further hypotheses and experimental tools can be developed.

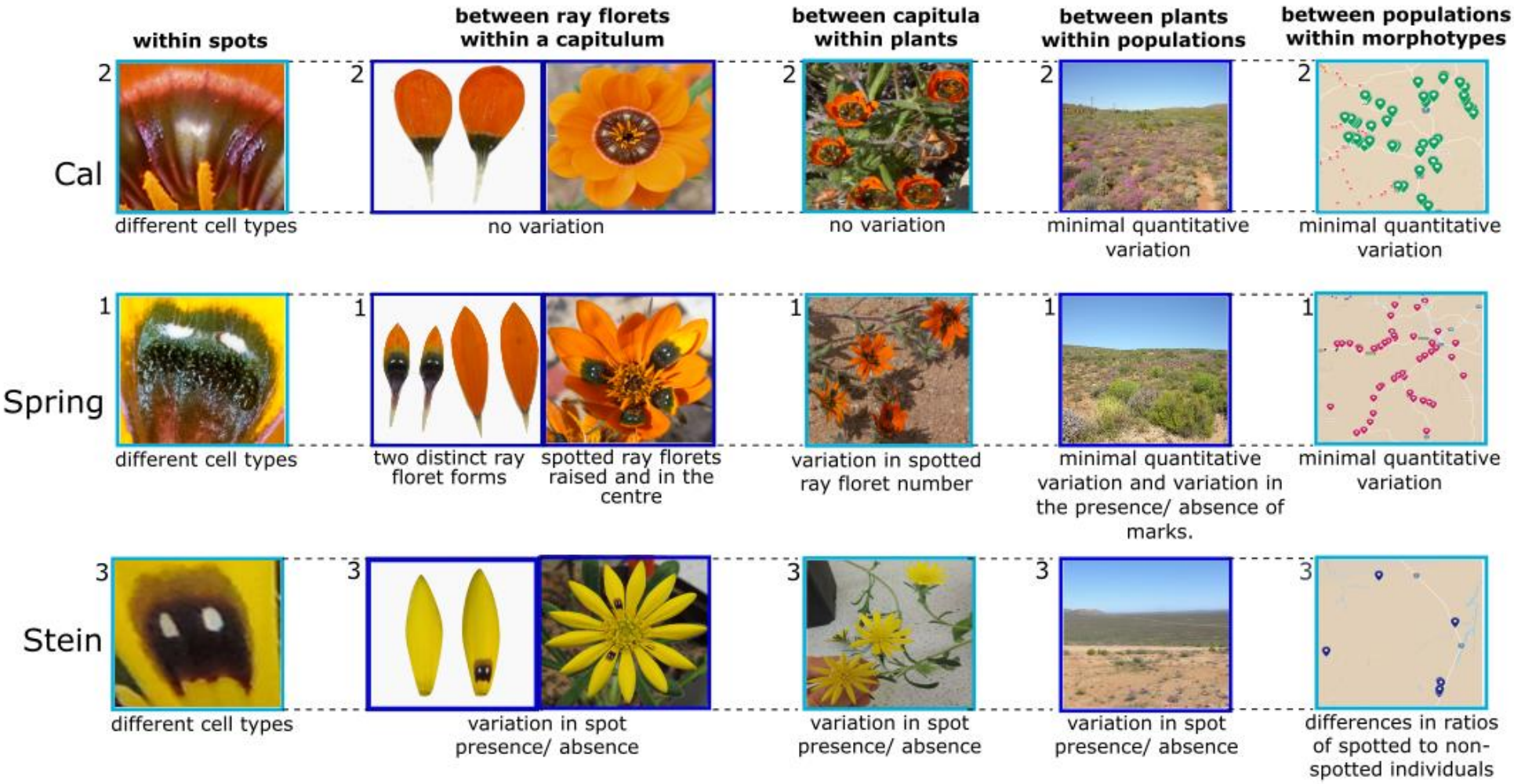


Fig 7.1. A representation of the variation present in floral phenotypes of *G. diffusa* across different scales. Variation is compared between Cal, Spring, and Stein morphotypes.

Bibliography

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. 2005. 'FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex.' *Science* 309(5737): 1052-1056.

Ahearn, K. P., Johnson, H. A., Weigel, D., and Wagner, D. R. 2001. 'NFL1, a *Nicotiana tabacum* LEAFY-like gene, controls meristem initiation and floral structure.' *Plant and Cell Physiology* 42(10): 1130–39.

Ahn, S. G., Liu, P. C., Klyachko, K., Morimoto, R. I., and Thiele, D. J. 2001. 'The loop domain of heat shock transcription factor 1 dictates DNA-binding specificity and responses to heat stress.' *Genes and Development* 15(16): 2134–45.

Airoldi, C. A., Hearn, T. J., Brockington, S. F., Webb, A. A., and Glover, B. J. 2019. 'TTG1 proteins regulate circadian activity as well as epidermal cell fate and pigmentation.' *Nature Plants* 5(11): 1145–53.

Albert, N. W., Lewis, D. H., Zhang, H., Schwinn, K. E., Jameson, P. E., and Davies, K. M. 2011a. 'Members of an R2R3-MYB transcription factor family in *Petunia* are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning.' *The Plant Journal* 65(5): 771–84.

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M. R., Ngo, H., Jameson, P. E., and Schwinn, K. E. 2014. 'A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots.' *The Plant Cell* 26(3): 962–80.

Albert, N. W., Butelli, E., Moss, S. M., Piazza, P., Waite, C. N., Schwinn, K. E., Davies, K. M., and Martin, C. 2020. 'Discrete bHLH transcription factors play functionally overlapping roles in pigmentation patterning in plowers of *Antirrhinum majus*.' *New Phytologist* (early release).

Albert, S., Delsey, M., and Devic, M. 1997. 'BANYULS, a novel negative regulator of flavonoid biosynthesis in *Arabidopsis* seed coat.' *The Plant Journal* 11(2): 289–99.

Alexander, D. H., Novembre, J., and Lange, K. 2009. 'Fast model-based estimation of ancestry in unrelated individuals.' *Genome Research* 19(9): 1655–64.

Almeida, J., Rocheta, M., and Galego, L. 1997. 'Genetic control of flower shape in *Antirrhinum majus*.' *Development* 124(7): 1387–92.

Alonso, C., Pérez, R., Bazaga, P., Medrano, M., and Herrera, C. M. 2018. 'Within-plant variation in seed size and inflorescence fecundity is associated with epigenetic mosaicism in the shrub *Lavandula latifolia* (Lamiaceae).' *Annals of Botany* 121(1): 153–60.

Ambrose, B. A., and Ferrándiz, C. 2018. 'Development and the evolution of plant form' *Annual Plant Reviews online*: 277-320.

Angenent, G. C., and Colombo, L. (1996). Molecular control of ovule development. *Trends in Plant Science*: 1(7), 228-232.

Armbruster, W. S., and Muchhala, N. 2009. 'Associations between floral specialization and species diversity: cause, effect, or correlation?' *Evolutionary Ecology* 23 (1): 159-179.

Armbruster, W. S. 2014. 'Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy.' *Annals of Botany Plants* 6: plu003.

An, C. H., Lee, K. W., Lee, S. H., Jeong, Y. J., Woo, S. G., Chun, H., Park, Y. I., Kwak, S. S., and Kim, C. Y. 2015. 'Heterologous expression of *IbMYB1a* by different promoters exhibits different patterns of anthocyanin accumulation in tobacco.' *Plant Physiology and Biochemistry* 89: 1–10.

Andersen, Ø. M., and Jordheim, M. 2006. 'The Anthocyanins.' In *In Ø. M. Andersen and K. R. Markham (Eds.), Flavonoids: Chemistry, Biochemistry and Applications*, Boca Raton: CRC Press, 471–553.

Andrews, S. 2010. 'FastQC: A quality control tool for high throughput sequence data [Online].' http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Angeloni, F., Joop Ouborg, N., and Leimu, R. 2011. 'Meta-analysis on the association of population size and life history with inbreeding depression in plants.' *Biological Conservation* 144(1): 35–43.

Ariel, F. D., Manavella, P. A., Dezar, C. A., and Chan, R. L. 2007. 'The true story of the HD-Zip family.' *Trends in Plant Science* 12(9): 419-426.

Asen, S. 1983. 'High pressure liquid chromatographic analysis of flavonoid in geranium florets as an adjunct for cultivar identification.' *Journal American Society for Horticultural Science* 108: 845–50.

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., Mayjonade, B., and Legrand, L. 2017. 'The sunflower genome provides insights into oil metabolism, flowering and asterid evolution.' *Nature* 546 (7656): 148–52.

Bai, Y., Pattanaik, S., Patra, B., Werkman, J. R., Xie, C. H., and Yuan, L. 2011. 'Flavonoid-related basic helix-loop-helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active.' *Planta* 234(2): 363–75.

Balamurali, G. S., Krishna, S., and Somanathan, H. 2015. 'Senses and signals: evolution of floral signals, pollinator sensory systems and the structure of plant–pollinator interactions.' *Current Science*: 1852–61.

Bao, S., Zhang, Z., Lian, Q., Sun, Q., and Zhang, R. 2019. 'Evolution and expression of genes encoding TCP transcription factors in *Solanum Tuberosum* reveal the involvement of StTCP23 in plant defence.' *BMC Genetics* 20(1): 1–15.

Barrier, M., Robichaux, R. H., and Purugganan, M. D. 2001. 'Accelerated regulatory gene evolution in an adaptive radiation.' *Proceedings of the National Academy of Sciences of the United States of America* 98(18): 10208–13.

Bastías, A., Yañez, M., Osorio, S., Arbona, V., Gómez-Cadenas, A., Fernie, A. R., and Casaretto, J. A. 2014. 'The transcription factor AREB1 regulates primary metabolic pathways in tomato fruits.' *Journal of Experimental Botany* 65(9): 2351–63.

Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. 'Fitting linear mixed-effects models using {lme4}.' *Journal of Statistical Software* 67(1): 1–48.

Baublis, A., Spomer, A., and Berber-Jimenez, M. D. 1994. 'Anthocyanin pigments: comparison of extract stability.' *Journal of Food Science* 59(6): 1219–21.

Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., Koes, R., Roberts, K., and Martin, C. 2007. 'Control of cell and petal morphogenesis by R2R3 MYB transcription factors.' *Development* 134(9): 1691–1701.

Behringer, C., and Schwechheimer, C. 2015. 'B-GATA transcription factors – insights into their structure, regulation, and role in plant development.' *Frontiers in Plant Science* 6: 1–12.

Bello, M. A., Álvarez, I., Torices, R., and Fuertes-Aguilar, J. 2013. 'Floral development and evolution of capitulum structure in *Anacyclus* (anthemideae, asteraceae).' *Annals of Botany* 112(8): 1597–1612.

Benlloch, R., Berbel, A., Serrano-Mislata, A., and Madueño, F. 2007. 'Floral initiation and inflorescence

architecture: a comparative view.' *Annals of Botany* 100(3): 659–76.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. 2012. 'Genbank.' *Nucleic Acids Research* 41: D36-D42.

Berbel, A., Navarro, C., Ferrándiz, C., Cañas, L. A., Madueño, F., and Beltrán, J. P. 2001. 'Analysis of PEAM4, the pea AP1 functional homologue, supports a model for AP1-like genes controlling both floral meristem and floral organ identity in different plant species.' *The Plant Journal* 25(4): 441–51.

Besnard, F., Refahi, Y., Morin, V., Marteaux, B., Brunoud, G., Chambrier, P., Rozier, F., Mirabet, V., Legrand, J., Lainé, S., and Thévenon, E. 2014. 'Cytokinin signalling inhibitory fields provide robustness to phyllotaxis.' *Nature* 505(7483): 417–21.

Binder, B. M., Walker, J. M., Gagne, J. M., Emborg, T. J., Hemmann, G., Bleecker, A. B., and Vierstra, R. D. 2007. 'The Arabidopsis EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling.' *The Plant Cell* 19(2): 509-523.

Blanc, G., and Wolfe, K. H. 2004. 'Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution.' *The Plant Cell* 16(7): 1679–91.

Blázquez, M. A., Ferrándiz, C., Madueño, F., and Parcy, F. 2006. 'How floral meristems are built.' *Plant Molecular Biology* 60: 855–70.

De Boer, J. G., and Dicke, M. 2004. 'The role of methyl salicylate in prey searching behavior of the predatory mite *Phytoseiulus persimilis*.' *Journal of Chemical Ecology* 30(2): 255–71.

Bolker, J. A. 2000. 'Modularity in development and why It matters to evo-devo.' *American Zoologist* 40(5): 770–76.

Bollhöner, B., Prestele, J., and Tuominen, H. 2012. 'Xylem cell death: emerging understanding of regulation and function.' *Journal of Experimental Botany* 63(3): 1081-1094.

Boucher, F. C., Verboom, G. A., Musker, S., and Ellis, A. G. 2017. 'Plant size: A key determinant of diversification?' *New Phytologist* 216(1): 24–31.

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. 1991. Genetic interactions among floral homeotic genes of Arabidopsis. *Development*, *112*(1), 1-20.

Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M. and Smyth, D. R. 1993. 'Control of flower development in *Arabidopsis thaliana* by APETALA 1 and interacting genes.' *Development* 119(3): 721–43.

Bradshaw Jr, H. D., and Schemske, D. W. 2003. 'Allele substitution at a flower colour locus produces a pollinatory shift in monkeyflowers.' *Nature* 426(6963): 176–78.

Bradshaw Jr, H. D., Wilbert, S. M., Otto, K. G., and Schemske, D. W. 1995. 'Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*).' *Nature* 376: 762–65.

Bradshaw, P., and Cowling, R. M. 2014. 'Landscapes, rock types, and climate of the Greater Cape Floristic Region.' In *Fynbos: Ecology, Evolution and Conservation of a Megadiverse Region*, 26–46. Oxford University Press.

Brewer, P. B., Howles, P. A., Dorian, K., Griffith, M. E., Ishida, T., Kaplan-Levy, R. N., Kilinc, A., and Smyth, D. R. 2004. 'PETAL LOSS, a trihelix transcription factor gene, regulates perianth architecture in the Arabidopsis flower.' *Development* 131(16): 4035-4045.

Britton, M. N., Hedderson, T. A., and Verboom, G. A. 2014. 'Topography as a driver of cryptic speciation

in the high-elevation Cape sedge *Tetraria triangularis* (Boeck.) C. B. Clarke (cyperaceae: schoeneae).' *Molecular Phylogenetics and Evolution* 77(1): 96–109.

Brody, A. K. 1992. 'Oviposition choices by a pre-dispersal seed predator (*Hylemya* Sp.) - I. Correspondence with hummingbird pollinators, and the role of plant size, density and floral morphology.' *Oecologia* 91(1): 56–62.

Brouillard, R. 1983. 'The in Vivo Expression of Anthocyanin Colour in Plants.' *Phytochemistry* 22(6): 1311–23.

Campbell, D. 1986. 'Predicting plant reproductive success from models of competition for pollination.' *Oikos* 47(3): 257–66.

Candolin, U. 2003. 'The use of multiple cues in mate choice.' *Biological Reviews of the Cambridge Philosophical Society* 78(4): 575–95.

Cardinal, S., and Danforth, B. N. 2013. 'Bees diversified in the age of eudicots.' *Proceedings of the Royal Society B: Biological Sciences* 280(1755).

Carey, C. C., Strahle, J. T., Selinger, D. A., and Chandler, V. L. 2004. 'Mutations in the *Pale Aleurone Color1* regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis thaliana*.' *The Plant Cell* 16(2): 450–64.

Carroll, S. B. 2008. 'Evo-Devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution.' *Cell* 134(1): 25–36.

Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-Ponce, B., Sánchez, M. D. L. P., Álvarez-Buylla, E. R., and Garay-Arroyo, A. 2019. 'MADS-Box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants.' *Frontiers in Plant Science* 10: 853.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. and Postlethwait, J. H. 2011. 'Stacks: Building and genotyping loci de novo from short-read sequences.' *G3: Genes, Genomes, Genetics* 1(3): 171–82.

Chanderbali, A. S., Berger, B. A., Howarth, D. G., Soltis, P. S. and Soltis, D. E. 2016. 'Evolving ideas on the origin and evolution of flowers: New Perspectives in the Genomic Era.' *Genetics* 202(4): 1255–65.

Chandler, J. W. 2011. 'Founder cell specification.' *Trends in Plant Science* 16(11): 607–13.

Chandler, J. W., Jacobs, B., Cole, M., Comelli, P., and Werr, W. 2011. 'DORNRÖSCHEN-LIKE expression marks *Arabidopsis* floral organ founder cells and precedes auxin response maxima.' *Plant Molecular Biology* 76(1–2): 171–85.

Chase, B. M., and Meadows, M. E. 2007. 'Late quaternary dynamics of Southern Africa's winter rainfall zone.' *Earth-Science Reviews* 84(3): 103–38.

Chaves-Silva, S., Dos Santos, A. L., Chalfun-Júnior, A., Zhao, J., Peres, L. E., and Benedito, V. A. 2018. 'Understanding the genetic regulation of anthocyanin biosynthesis in plants – tools for breeding purple varieties of fruits and vegetables.' *Phytochemistry* 153: 11–27.

Chen, D., Pan, X., Xiao, P., Farwell, M. A. and Zhang, B. 2011. 'Evaluation and identification of reliable reference genes for pharmacogenomics, toxicogenomics, and small RNA expression analysis.' *Journal of Cellular Physiology* 226(10): 2469–77.

Cheng, X., Xiong, R., Yan, H., Gao, Y., Liu, H., Wu, M. and Xiang, Y. 2019. 'The trihelix family of

transcription factors: functional and evolutionary analysis in moso bamboo (*Phyllostachys edulis*).' *BMC Plant Biology* 19(1): 1–20.

Chow, C. N., Lee, T. Y., Hung, Y. C., Li, G. Z., Tseng, K. C., Liu, Y. H., Kuo, P. L., Zheng, H. Q. and Chang, W. C 2019. 'Plantpan3.0: A new and updated resource for reconstructing transcriptional regulatory networks from chip-seq experiments in plants.' *Nucleic Acids Research* 47(D1): D1155–63.

Chu, C. C., Qu, N., Bassüner, B., and Bauwe, H. 1997. 'Genetic transformation of the C3-C4 intermediate plant, *Flaveria pubescens* (Asteraceae).' *Plant Cell Reports* 16 (10): 715–18.

Coberly, L. C., and Rausher, M. D. 2003. 'Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: amelioration of heat stress.' *Molecular Ecology* 12(5): 1113–24.

Coberly, L. C., and Rausher, M. D. 2008. 'Pleiotropic effects of an allele producing white flowers in *Ipomoea purpurea*.' *Evolution* 62(5): 1076–85.

Coen, E. S., and Meyerowitz, E. M. 1991. 'The war of the whorls: genetic interactions controlling flower development.' *Nature* 353(6339): 31.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. 2008. 'Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning.' *Nature* 452(7184): 215–19.

Colasanti, J., Yuan, Z., and Sundaresan, V. 1998. 'The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize.' *Cell* 93(4): 593-603.

Coleman, G. D., and Ernst, S. G. 1989. 'In vitro shoot regeneration of *Populus deltoides*: effect of cytokinin and genotype.' *Plant Cell Reports* 8: 459–62.

Cominelli, E., and Tonelli, C. 2009. 'A new role for plant R2R3-MYB transcription factors in cell cycle regulation.' *Cell Research* 19(11): 1231-1232.

Cone, K. C., Burr, F. A., and Burr, B. 1986. 'Molecular analysis of the maize anthocyanin regulatory locus C1.' *Proceedings of the National Academy of Sciences of the United States of America* 83(24): 9631–35.

Cooley, A. M., and Willis, J. H., 2009. 'Genetic divergence causes parallel evolution of flower color in Chilean *Mimulus*.' *New Phytologist* 183(3): 729-739.

Cooley, A. M., Modliszewski, J. L., Rommel, M. L., and Willis, J. H. 2011. 'Gene duplication in *Mimulus* underlies parallel floral evolution via independent trans-regulatory changes.' *Current Biology* 21(8): 700–704.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., and Coupland, G. 2007. 'FT Protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*.' *Science* 316(5827): 1030–33.

Corradini, E., Foglia, P., Giansanti, P., Gubbiotti, R., Samperi, R., and Lagana, A., 2011. 'Flavonoids: chemical properties and analytical methodologies of identification and quantitation in foods and plants.' *Natural Product Research* 25(5): 469–95.

Ćosić, T., Motyka, V., Raspor, M., Savić, J., Cingel, A., Vinterhalter, B., Vinterhalter, D., Trávníčková, A., Dobrev, P. I., Bohanec, B., and Ninković, S. 2015. '*In vitro* shoot organogenesis and comparative analysis of endogenous phytohormones in kohlrabi (*Brassica oleracea* var. gongylodes): Effects

of genotype, explant type and applied cytokinins.' *Plant Cell, Tissue and Organ Culture* 121(3): 741–60.

Cowling, R. M., Esler, K. J., and Rundel, P. W. 1999. 'Namaqualand, South Africa - An overview of a unique winter-rainfall desert ecosystem.' *Plant Ecology* 142(1–2): 3–21.

Cowling, R. 2015. *Namaqualand: A Succulent Desert.* South Africa: Penguin Random House.

Cowling, R. M., and Hilton-Taylor, C. 2001. 'Plant biogeography, endemism and diversity.' In *The Karoo – Ecological Patterns and Processes*, Cambridge: Cambridge University Press, 42–56.

Cowling, R. M., Procheş, Ş., and Partridge, T. C. 2009. 'Explaining the uniqueness of the Cape flora: incorporating geomorphic evolution as a factor for explaining its diversification.' *Molecular Phylogenetics and Evolution* 51(1): 64–74.

Cowling, R. M., Rundel, P. W., Desmet, P. G., and Esler, K. J. 1998. 'Extraordinary high regional-scale plant diversity in southern African arid lands : subcontinental and global.' *Diversity and Distributions* 4(1): 27–36.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., and McVean, G. 2011. 'The variant call format and VCFtools.' *Bioinformatics* 27(15): 2156–58.

Darwin, C. 1862. *On the various contrivances by which British and foreign orchids are fertilized*. London, UK: John Murray.

Davies, K. M., Albert, N. W., and Schwinn, K. E. 2012. 'From landing lights to mimicry: the molecular regulation of flower colouration and mechanisms for pigmentation patterning.' *Functional Plant Biology* 39(8): 619–38.

Davies, T. J., Barraclough, T. G., Chase, M. W., Soltis, P. S., Soltis, D. E., and Savolainen, V. 2004. 'Darwin's abominable mystery: insights from a supertree of the angiosperms.' *Proceedings of the National Academy of Sciences of the United States of America* 101(7): 1904–9.

Davis, T. L. 2020. 'Optparse: command line option parser.' https://cran.rproject.org/package=optparse.

Day, T. 2000. 'Sexual selection and the evolution of costly female preferences: spatial effects.' *Evolution* 54(3): 715–30.

De Luca, P. A., and Vallejo-Marin, M., 2013. 'What's the 'buzz' about? The ecology and evolutionary significance of buzz-pollination.' *Current Opinion in Plant Biology*, 16(4):429-435.

Dek, M. S. P., Padmanabhan, P., Sherif, S., and Subramanian, J. 2017. 'Upregulation of phosphatidylinositol 3-kinase (Pi3k) enhances ethylene biosynthesis and accelerates flower senescence in transgenic *Nicotiana tabacum* L.' *International Journal of Molecular Sciences* 18(7): 1–15.

Delgado-Vargas, F., Jiménez, A. R., Paredes-López, O., and Francis, F. J. 2000. 'Natural pigments: carotenoids, anthocyanins, and betalains—characteristics, biosynthesis, processing, and stability.' *Critical Reviews in Food Science and Nutrition* 40(3): 173-289.

Deng, C., Li, S., Feng, C., Hong, Y., Huang, H., Wang, J., Wang, L., and Dai, S. 2019. 'Metabolite and gene expression analysis reveal the molecular mechanism for petal colour variation in six *Centaurea cyanus* cultivars.' *Plant Physiology and Biochemistry* 142: 22–33.

Deng, W., Pu, X. A., Goodman, R. N., Gordon, M. P., and Nester, E. W., 1995. 'T-DNA genes responsible for inducing a necrotic response on grape vines.' *Molecular Plant-Microbe Interactions* 8(4): 538–

48.

Deng, X., Yang, J., Wu, X., Li, Y., and Fei, X. 2014. 'A C2H2 zinc finger protein FEMU2 is required for Fox1 expression in *Chlamydomonas reinhardtii*.' *PLoS ONE* 9(12): 1–24.

Deng, Y., and Lu., S. 2017. 'Biosynthesis and regulation of phenylpropanoids in plants.' *Critical Reviews in Plant Sciences* 36(4): 257–90.

Des Marais, D. L., and Rausher, M. D. 2008. 'Escape from adaptive conflict after duplication in an anthocyanin pathway gene.' *Nature Letters* 454(7205): 762–65.

Desmet, P. G. 2007. 'Namaqualand - A brief overview of the physical and floristic environment.' *Journal of Arid Environments* 70(4): 570–87.

Desmet, P. G., and Cowling, R. M. 1999. 'Biodiversity, habitat and range-size aspects of a flora from a winter-rainfall desert in north-western Namaqualand, South Africa.' *Plant Ecology* 142(1–2): 23–33.

Dias, A. P., Braun, E. L., McMullen, M. D., and Grotewold, E. 2003. 'Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication.' *Plant Physiology* 131(2): 610–20.

Dick, C. W., Hardy, O. J., Jones, F. A., and Petit, R. J. 2008. 'Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees.' *Tropical Plant Biology* 1(1): 20–33.

Dieckmann, U., and Doebeli. M. 1999. 'On the origin of species by sympatric speciation.' *Nature* 400(6742): 354–57.

Diester-Haass, L., Meyers, P. A., and Vidal, L. 2002. 'The late Miocene onset of high productivity in the Benguela current upwelling system as part of a global pattern.' *Marine Geology* 180(1–4): 87–103.

Ding, B., Patterson, E. L., Holalu, S. V., Li, J., Johnson, G. A., Stanley, L. E., Greenlee, A. B., Peng, F., Bradshaw, H. D., Blackman, B. K., and Yuan, Y. W. 2018. 'Formation of periodic pigment spots by the reaction-diffusion mechanism.' *BioRxiv*: 403600.

Ding, B., and Yuan. Y. W. 2016. 'Testing the utility of fluorescent proteins in *Mimulus lewisii* by an *Agrobacterium*-mediated transient assay.' *Plant Cell Reports* 35(4): 771–77.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M. F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Current Biology* 14(21): 1935-1940.

Divya, K., Kishor, P. K., Bhatnagar-Mathur, P., Singam, P., Sharma, K. K., Vadez, V., and Reddy, P. S., 2019. 'Isolation and functional characterization of three abiotic stress-inducible (Apx, Dhn and Hsc70) promoters from pearl millet (*Pennisetum glaucum* L.).' *Molecular Biology Reports* 46(6): 6039-6052.

Dodson, G., and Wlodawer, A. 1998. 'Catalytic triads and their relatives.' *Trends in Biochemical Sciences* 23(9): 347–52.

Doebeli, M., and Dieckmann, U. 2003. 'Interim report: speciation along environmental gradients.' *Nature* 421: 259–264.

Dong, G., Ni, Z., Yao, Y., Nie, X., and Sun, Q. 2007. 'Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development.' *Plant Molecular Biology* 63(1): 73-84.

Dray, S., and Dufour, A. B. 2007. 'The Ade4 package: implementing the duality diagram for ecologists.'

Journal of Statistical Software 22(4): 1–20.

Du, H., Liang, Z., Zhao, S., Nan, M. G., Tran, L. S. P., Lu, K., Huang, Y. B. and Li, J. N. 2015. 'The evolutionary history of R2R3-Myb proteins across 50 eukaryotes: new insights into subfamily classification and expansion.' *Scientific Reports* 5: 1–7.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. 2010. 'MYB transcription factors in *Arabidopsis*.' *Trends in Plant Science* 15(10): 573–81.

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G. and Petit, R. J. 2007. 'Can population genetic structure by predicted from life-history traits?' *The American Naturalist* 169(5): 662–72.

Duminil, J., Hardy, O. J., and Petit, R. J. 2009. 'Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure.' *BMC Evolutionary Biology* 9(1): 1–14.

Duncan, G., and Ellis, A. G. 2011. '723. *Gorteria diffusa*.' *Curtis's Botanical Magazine* 28(4): 341–48.

Durbin, M. L., Lundy, K. E., Morrell, P. L., Torres-Martinez, C. L., and Clegg, M. T. 2003. 'Genes that determine lower color: the role of regulatory changes in the evolution of phenotypic adaptations.' *Molecular Phylogenetics and Evolution* 29(3): 507–18.

Dyer, A. G., Whitney, H. M., Arnold, S. E., Glover, B. J., and Chittka, L. 2006. 'Bees associate warmth with floral colour.' *Nature* 442(7102): 525.

Eckhart, V. M., Rushing, N. S., Hart, G. M., and Hansen, J.D. 2006. 'Frequency-dependent pollinator foraging in polymorphic *Clarkia xantiana* ssp. *xantiana* populations: implications for flower colour evolution and pollinator interactions.' *Oikos* 112(2): 412–21.

Eklund, D. M., Ståldal, V., Valsecchi, I., Cierlik, I., Eriksson, C., Hiratsu, K., Ohme-Takagi, M., Sundström, J. F., Thelander, M., Ezcurra, I., and Sundberg, E. 2010. 'The Arabidopsis thaliana STYLISH1 protein acts as a transcriptional activator regulating auxin biosynthesis.' *The Plant Cell* 22(2): 349-363.

Ellis, A. G., and Weis, A. E. 2006. 'Coexistence and Differentiation of 'Flowering Stones': The Role of Local Adaptation to Soil Microenvironment.' *Journal of Ecology* 94(2): 322–35.

Ellis, A. G., Weis, A. E., and Gaut. B. S. 2006. 'Evolutionary radiation of 'stone plants' in the genus *Argyroderma* (Aizoaceae): Unraveling the effects of landscape, habitat, and flowering time.' *Evolution* 60(1): 39.

Ellis, A. G., Verboom, G. A., van der Niet, T., Johnson, S. D., Linder, H. P., Allsopp, N., and Colville, J. F. 2013. 'Speciation and extinction in the Greater Cape Floristic Region.' *Etica e Politica* 15(1): 583–605.

Ellis, A. G., Brockington, S. F., de Jager, M. L., Mellers, G., Walker, R. H., and Glover, B. J. 2014. 'Floral trait variation and integration as a function of sexual deception in *Gorteria diffusa*.' *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 369: 1–13.

Ellis, A. G., and Johnson, S. D. 2009. 'The evolution of floral variation without pollinator shifts in *Gorteria diffusa* (Asteraceae).' *American Journal of Botany* 96(4): 793–801.

Ellis, A. G., and Johnson, S. D. 2010. 'Floral mimicry enhances pollen export: the evolution of pollination by sexual deceit outside of the Orchidaceae.' *The American Naturalist* 176(5): E143–51.

Ellis, A. G., and Johnson, S. D. 2012. 'Lack of floral constancy by bee fly pollinators: implications for ethological isolation in an African daisy.' *Behavioral Ecology* 23(4): 729–34.

Ellis, A. G., Weis, A. E., and Gaut, B. S. 2007. 'Spatial scale of local adaptation and population genetic structure in a miniature succulent, *Argyroderma pearsonii*.' *New Phytologist* 174: 904–14.

Ellstrand, N. C. 2014. 'Is gene flow the most important evolutionary force in plants?' *American Journal of Botany* 101(5): 737–53.

Elomaa, P., Honkanen, J., Puska, R., Seppänen, P., Helariutta, Y., Mehto, M., Kotilainen, M., Nevalainen, L., and Teeri, T.H. 1993. '*Agrobacterium*-mediated transfer of antisense *chalcone synthase* cDNA to *Gerbera hybrida* inhibits flower pigmentation.' *Biotechnology* 11(4): 508–11.

Elomaa, P., Uimari, A., Mehto, M., Albert, V. A., Laitinen, R. A., and Teeri, T. H. 2003. 'Activation of anthocyanin biosynthesis in *Gerbera hybrida* (Asteraceae) suggests conserved protein-protein and protein-promoter interactions between the anciently diverged monocots and eudicots.' *Plant Physiology* 133(4): 1831–42.

Eltz, T., Roubik, D. W., and Lunau, K. 2005. 'Experience-dependent choices ensure species-specific fragrance accumulation in male orchid bees.' *Behavioral Ecology and Sociobiology* 59(1): 149–56.

Endler, J. A. 1992. 'Signals, signal onditions, and the direction of evolution.' *The American Naturalist* 139: 125–53.

Endress, P. K. 2012. 'The immense diversity of floral monosymmetry and asymmetry across angiosperms.' *Botanical Review* 78(4): 345–97.

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. 2007. 'Red Colouration in Apple Fruit Is Due to the Activity of the MYB Transcription Factor, MdMYB10.' *The Plant Journal* 49(3): 414–27.

Fang, Y., Xie, K., Hou, X., Hu, H., and Xiong, L. 2010. 'Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses.' *Molecular Genetics and Genomics* 283(2): 157-169.

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., Prinsi, B., Jaarsma, R., Tarhan, E., de Boer, A. H., and Di Sansebastiano, G. P. 2014. 'Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color.' *Cell Reports* 6(1): 32–43.

Fattorini, R., and Glover. B. J. 2020. 'Molecular mechanisms of pollination biology.' *Annual Review of Plant Biology* 71(1): 1–29.

Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., and Nosil, P. 2013. 'Geographic mode of speciation and genomic divergence.' *Annual Review of Ecology, Evolution, and Systematics* 44: 73–97.

Fehr, C., and Rausher, M. D. 2004. 'Effects of variation at the flower-colour a locus on mating system parameters in *Ipomoea purpurea*.' *Molecular Ecology* 13(7): 1839–47.

Feller, A., MacHemer, K., Braun, E. L., and Grotewold, E. 2011. 'Evolutionary and comparative analysis of MYB and bHLH plant transcription factors.' *The Plant Journal* 66(1): 94–116.

Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., and Thomson, J. D. 2004. 'Pollination syndromes and floral specialization.' *Annual Review of Ecology, Evolution, and Systematics* 35: 375–403.

Fernández Gómez, J., and Wilson, Z.A. 2014. 'A barley PHD finger transcription factor that confers male sterility by affecting tapetal development.' *Plant Biotechnology Journal* 12(6): 765-777.

Fernández-Mazuecos, M., and Glover, B. J. 2017. 'The evo-devo of plant speciation.' *Nature Ecology and Evolution* 1(4): 1–9.

Ferrer, J. L., Austin, M. B., Stewart, C., and Noel, J. P. 2008. 'Structure and function of enzymes involved in the biosynthesis of phenylpropanoids.' *Plant Physiology and Biochemistry* 46(3): 356–70.

Figueiredo, P., George, F., Tatsuzawa, F., Toki, K., Saito, N., and Brouillard, R. 1999. 'New features of intramolecular copigmentation by acylated anthocyanins.' *Phytochemistry* 51(1): 125–32.

Finnegan, E. J., Peacock, W. J., and Dennis, E. S. 2000. 'DNA methylation, a key regulator of plant development and other processes.' *Current Opinion in Genetics and Development* 10(2): 217–23.

Fischer, T.C., Halbwirth, H., Meisel, B., Stich, K., and Forkmann, G. 2003. 'Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from *Malus domestica* and *Pyrus communis* Cultivars and the consequences for flavonoid metabolism.' *Archives of Biochemistry and Biophysics* 412(2): 223–30.

Fitzpatrick, B. M., Fordyce, J. A., and Gavrilets, S. 2009. 'Pattern, process and geographic modes of speciation.' *Journal of Evolutionary Biology* 22(11): 2342–47.

Flatscher, R., Frajman, B., Schönswetter, P., and Paun, O. 2012. 'Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation?' *Genetics Research International* 2012: 698421–9.

Forest, F., Goldblatt, P., Manning, J. C., Baker, D., Colville, J. F., Devey, D. S., Jose, S., Kaye, M., and Buerki, S. 2014. 'Pollinator shifts as triggers of speciation in painted petal Irises (*Lapeirousia iridaceae*).' *Annals of Botany* 113(2): 357–71.

Fornara, F., Panigrahi, K. C., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J. A., and Coupland, G. 2009. '*Arabidopsis* DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response.' *Developmental Cell* 17(1): 75-86.

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. 2014. 'DNA-binding specificities of plant transcription factors and their potential to define target genes.' *Proceedings of the National Academy of Sciences of the United States of America* 111(6): 2367–72.

Fu, Z., Jiang, H., Chao, Y., Dong, X., Yuan, X., Wang, L., Zhang, J., Xu, M., Wang, H., Li, Y., and Gao, J. 2020. 'Three Paralogous R2R3-MYB Genes Contribute to Delphinidin-Related Anthocyanins Synthesis in *Petunia Hybrida*.' *Journal of Plant Growth Regulation* 1-14.

Fukada-Tanaka, S., Inagaki, Y., Yamaguchi, T., Saito, N., and Iida, S. 2000. 'Colour-enhancing protein in blue petals.' *Nature* 407(6804): 581.

Funk, D. J., Nosil, P., and Etges, W. J. 2006. 'Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa.' *Proceedings of the National Academy of Sciences of the United States of America* 103(9): 3209–13.

Funk, V. A., and Chan, R. 2008. 'Phylogeny of the spiny African daisies (Compositae, tribe Arctotideae, subtribe Gorteriinae) based on TrnL-F, NdhF, and ITS sequence data.' *Molecular Phylogenetics and Evolution* 48(1): 47–60.

Galen, C. 1999. 'Why do flowers vary? The functional ecology of variation in flower size and form within natural plant populations.' *Bioscience* 49(8): 631–60.

Gamba, D., and Muchhala, N. 2020. 'Global patterns of population genetic differentiation in seed plants.' *Molecular Ecology* 29(18): 3413–28.

Ganger, M. T., Dietz, G. D., and Ewing, S. J. 2017. 'A common base method for analysis of qPCR data and the application of simple blocking in qPCR experiments.' *BMC Bioinformatics* 18(1): 1–11.

Garcês, H. M. P., Spencer, V. M. R., and Kim, M. 2016. 'Control of floret symmetry by RAY3, SvDIV1B, and SvRAD in the capitulum of *Senecio vulgaris*.' *Plant Physiology* 171(3): 2055–68.

Garfield, D. A., Runcie, D. E., Babbitt, C. C., Haygood, R., Nielsen, W. J. and Wray, G. A. 2013. 'The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network.' *PLOS Biology* 11(10): e1001696.

Garfinkel, D. J., and Nester, E. W. 1980. '*Agrobacterium tumefaciens* mutants affected in crown gall tumorigenesis and octopine catabolism.' *Journal of Bacteriology* 144(2): 732–43.

Gaskett, A. C. 2011. 'Orchid pollination by sexual deception: pollinator perspectives.' *Biological Reviews* 86(1): 33–75.

Gauthier, J., de Silva, D.L., Gompert, Z., Whibley, A., Houssin, C., Le Poul, Y., McClure, M., Lemaitre, C., Legeai, F., Mallet, J., and Elias, M. 2020. 'Contrasting genomic and phenotypic outcomes of hybridization between pairs of mimetic butterfly taxa across a suture zone.' *Molecular Ecology* 29(7): 1328–43.

van Geldermalsen, M. 2016. 'An overview of the yeast one-hybrid assay.' *BiteSize Bio*. https://bitesizebio.com/25900/an-overview-of-the-yeast-one-hybrid-assay/.

Gendron, J.M., Pruneda-Paz, J.L., Doherty, C.J., Gross, A.M., Kang, S.E., and Kay, S.A. 2012. '*Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor.' *Proceedings of the National Academy of Sciences of the United States of America* 109(8): 3167–72.

Gerats, A. G. M., De Vlaming, P., Doodeman, M. A. B., and Schram, A. W. 1982. 'Genetic control of the conversion of dihydroflavonols into flavonols and anthocyanins in flowers of *Petunia hybrida*.' *Planta* 155(4): 364–68.

Gerats, A. G.M., Vrijlandt, E., Wallroth, M., and Schram, A. W. 1985. 'The influence of the genes An1, An2, and An4 on the activity of the enzyme UDP-glucose: flavonoid 3-O-glucosyltransferase in flowers of *Petunia hybrida*.' *Biochemical Genetics* 23(7–8): 591–98.

Gerats, A. G., Farcy, E., Wallroth, M., Groot, S. P., and Schram, A. 1984. 'Control of anthocyanin synthesis in *Petunia hybrida* by multiple allelic series of the genes *AN1* and *AN2*.' *Genetics* 106(3): 501–8.

Givnish, T. J. 1997. 'Adaptive radiation and molecular systematics: aims and conceptual issues.' In *Molecular Evolution and Adaptive Radiation*, New York, NY, USA: Cambridge University Press, 1–54.

Givnish, T. J., Millam, K. C., Mast, A. R., Paterson, T. B., Theim, T. J., Hipp, A. L., Henss, J. M., Smith, J. F., Wood, K. R., and Sytsma, K. J. 2009. 'Origin, adaptive radiation and diversification of the Hawaiian Lobeliads (Sterales: Campanulaceae).' *Proceedings of the Royal Society B: Biological Sciences* 276(1656): 407–16.

Givnish, T. J. 2010. 'Ecology of plant speciation.' *Taxon* 59(5): 1329–66.

Givnish, T. J. 2015. 'Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution.' *New Phytologist* 207(2): 297–303.

Glover, B. J., Perez-Rodriguez, M., and Martin, C. 1998. 'Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor.' *Development* 125(17): 3497–3508.

Gómez, J. M. 2003. 'Herbivory reduces the strength of pollinator-mediated selection in the Mediterranean herb *Erysimum mediohispanicum*: consequences for plant specialization.' *American Naturalist* 162(2): 242–56.

Gómez, J. M., Perfectti, F., and Camacho, J. P. M. 2006. 'Natural selection on *Erysimum mediohispanicum* flower shape: insights into the evolution of zygomorphy.' *American Naturalist* 168(4): 531–45.

Gong, S., Ding, Y., Hu, S., Ding, L., Chen, Z., and Zhu, C. 2019. The role of HD-Zip class I transcription factors in plant response to abiotic stresses. *Physiologia Plantarum* 167(4): 516-525.

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. 2008. 'Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings.' *The Plant Journal* 53(5): 814–27.

Goodrich, J., Carpenter, R., and Coen, E. 1992. 'A common gene regulates pigmentation pattern in diverse plant species.' *Cell* 68(5): 955–64.

Gosselin, T. 2020. 'Comparisons of Fst computations.' https://thierrygosselin.github.io/assigner/articles/web_only/fst_comparisons.html.

Goudet, J. 2005. 'HIERFSTAT, a package for R to compute and test hierarchical F-statistics.' *Molecular Ecology Notes* 5(1): 184–86.

van der Graaff, E., Laux, T., and Rensing, S. A. 2009. 'The WUS homeobox-containing (WOX) protein family.' *Genome Biology* 10(12): 1-9.

Gramzow, L., Ritz, M. S., and Theißen, G. 2010. 'On the origin of MADS-domain transcription factors.' *Trends in Genetics* 26(4): 149–53.

Grant, V., and Grant, K. A. 1965. *Flower Pollination in the Phlox Family*. New York: Columbia University Press.

Grant, V. 1949. 'Pollination systems as isolating mechanisms in angiosperms.' *Evolution* 3(1): 82–97.

Griesbach, R. J., Asen, S., and Leonnarat, B. A. 1991. '*Petunia hybrida* anthocyanins acylated with caffeic acid.' *Phytochemistry* 30(5): 1729–31.

Grotewold, E. 2005. 'Plant metabolic diversity: a regulatory perspective.' *Trends in Plant Science* 10(2): 57–62.

Grotewold, E. 2006. 'The genetics and biochemistry of floral pigments.' *Annual Review of Plant Biology* 57(1): 761–80.

Guo, X., Zhang, Y., Tu, Y., Wang, Y., Cheng, W., and Yang, Y., 2018. 'Overexpression of an EIN3-binding F-box protein2-like gene caused elongated fruit shape and delayed fruit development and ripening in tomato.' *Plant Science* 272: 131-141.

Haberer, G., Hindemitt, T., Meyers, B. C., and Mayer, K. E. X. 2004. 'Transcriptional similarities, dissimilarities, and conservation of *cis*-elements in duplicated genes of *Arabidopsis*.' *Plant Physiology* 136(2): 3009–22.

Habu, Y., Hisatomi, Y., and Lida, S. 1998. 'Molecular characterization of the mutable flaked allele for flower variegation in the common Morning Glory.' *The Plant Journal* 16(3): 371–76.

Hanoteaux, S., Tielbörger, K., and Seifan, M. 2013. 'Effects of spatial patterns on the pollination success of a less attractive species.' *Oikos* 122(6): 867–80.

Hansen, G. 2000. 'Evidence for *Agrobacterium*-induced apoptosis in maize cells.' *Molecular Plant-Microbe Interactions* 13(6): 649–57.

Harborne, J. B. 1963. 'Plant polyphenols - XI. The tructure of acylated anthocyanins.' *Phytochemistry* 3(2): 151–60.

Harder, L. D., Strelin, M. M., Clocher, I. C., Kulbaba, M. W., and Aizen, M. A. 2019. 'The dynamic mosaic phenotypes of flowering plants.' *New Phytologist* 224(3): 1021–34.

Hartmann, H. E. K. 1991. 'Mesembryanthema.' *Contribution from the Bolus Herbarium.* 13: 75–157.

Hatlestad, G. J., Akhavan, N. A., Sunnadeniya, R. M., Elam, L., Cargile, S., Hembd, A., Gonzalez, A., McGrath, J. M., and Lloyd, A. M. 2015. 'The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway.' *Nature Genetics* 47(1): 92–96.

He, B., Shi, P., Lv, Y., Gao, Z., and Chen, G. 2020. 'Gene coexpression network analysis reveals the role of SRS genes in senescence leaf of maize (*Zea mays* L.).' *Journal of Genetics* 99(1): 1-10.

He, H., Ke, H., Keting, H., Qiaoyan, X., and Silan, D. 2013. 'Flower colour modification of *Chrysanthemum* by suppression of F3'H and overexpression of the exogenous *Senecio cruentus F3'5'H* gene.' *PLoS ONE* 8(11): 1–12.

Heinz, S. K., Mazzucco, R., and Dieckmann, U. 2009. 'Speciation and the evolution of dispersal along environmental gradients.' *Evolutionary Ecology* 23(1): 53–70.

Helariutta, Y., Elomaa, P., Kotilainen, M., Griesbach, R. J., Schröder, J., and Teeri, T. H. 1995. 'Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (Asteraceae).' *Plant Molecular Biology* 28(1): 47–60.

Helariutta, Y., Elomaa, P., Kotilainen, M., Seppänen, P., and Teeri, T.H., 1993. 'Cloning of cDNA coding for dihydroflavonol-4-reductase (DFR) and characterization of Dfr expression in the corollas of *Gerbera hybrida* var. Regina (Compositae).' *Plant Molecular Biology* 22(2): 183–93.

Henning, J. A., Peng, Y., Montague, M. A., and Teuber, L. R. 1992. 'Honey bee (Hymenoptera: Apidae) behavioral response to primary alfalfa (Rosales: Fabaceae) floral volatiles.' *Journal of Economic Entomology* 85(1): 233–39.

Hernández, I., Alegre, L., Van Breusegem, F., and Munné-Bosch, S. 2009. 'How relevant are flavonoids as antioxidants in plants?' *Trends in Plant Science* 14(3): 125–32.

Hernandez, J., Heine, G., Irani, N. G., Feller, A. and Kim, M. G. 2004. 'Mechanisms of cooperation between MYB and HLH transcription factors in the regulation of anthocyanin pigmentation.' *Journal of Biological Chemistry* (279): 48205–13.

Herrera, C. M., Medrano, M., Perez, R., Bazaga, P., and Alonso, C. 2019. 'Within-plant heterogeneity in fecundity and herbivory induced by localized DNA hypomethylation in the perennial herb *Helleborus foetidus*.' *American Journal of Botany* 106(6): 798–806.

Herrera, C. M., Bazaga, P., Pérez, R., and Alonso, C. 2020. 'Lifetime genealogical divergence within plants leads to epigenetic mosaicism in the long-lived shrub *Lavandula latifolia* (Lamiaceae).' *bioRxiv*: 1–34.

Himmelbach, A., Hoffmann, T., Leube, M., Höhener, B., and Grill, E. 2002. 'Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in *Arabidopsis*.' *The EMBO Journal* 21(12): 3029-3038.

Hoballah, M. E., Stuurman, J., Turlings, T. C., Guerin, P. M., Connetable, S., and Kuhlemeier, C 2005.

'The composition and timing of flower odour emission by wild *Petunia axillaris* coincide with the antennal perception and nocturnal activity of the pollinator *Manduca sexta*.' *Planta* 222(1): 141–50.

Hoekstra, H. E., and Coyne, J. A. 2007. 'The locus of evolution: evo devo and the genetics of adaptation.' *Evolution* 61(5): 995–1016.

Hong, Y., Tang, X., Huang, H., Zhang, Y. and Dai, S. 2015. 'Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in *Chrysanthemum*.' *BMC Genomics* 16(1): 1–18.

Hooykaas, P. J. J., Hofker, M., den Dulk-Ras, H., and Schilperoort, R. A. 1984. 'A comparison of virulence determinants in an octopine Ti plasmid , a nopaline Ti plasmid , and an Ri plasmid by complementation analysis of *Agrobacterium tumefaciens* mutants.' *Plasmid* 11: 195–205.

Horikoshi, M., and Tang, Y. 2018. 'Ggfortify: data visualization tools for statistical analysis results.' https://cran.r-project.org/package=ggfortify.

Horsch, R. B., Klee, H. J., Stachel, S., Winans, S. C., Nester, E. W., Rogers, S. G., and Fraley, R. T. 1986. 'Analysis of *Agrobacterium tumefaciens* virulence mutants in leaf discs.' *Proceedings of the National Academy of Sciences of the United States of America* 83(8): 2571–75.

Hothorn, T., Bretz, F., and Westfall, P. 2008. 'Simultaneous inference in general parametric models.' *Biometrical Journal* 50(3): 346–63.

Hsu, Y. H., Tagami, T., Matsunaga, K., Okuyama, M., Suzuki, T., Noda, N., Suzuki, M., and Shimura, H. 2017. 'Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase Involved in anthocyanin modification , a key enzyme determining blue coloration in *Lobelia erinus*.' *The Plant Journal* 89(2): 325–37.

Hu, W., and Ma, H. 2006. 'Characterization of a novel putative zinc finger gene MIF1: involvement in multiple hormonal regulation of *Arabidopsis* development.' *The Plant Journal* 45(3): 399-422.

Hu, S., Dilcher, D. L., Jarzen, D. M., and Winship Taylor, D. 2008. 'Early steps of angiosperm pollinator coevolution.' *Proceedings of the National Academy of Sciences* 105(1): 240–45.

Hu, W., Zuo, J., Hou, X., Yan, Y., Wei, Y., Liu, J., Li, M., Xu, B., and Jin, Z., 2015. The auxin response factor gene family in banana: genome-wide identification and expression analyses during development, ripening, and abiotic stress. *Frontiers in Plant Science* 6: 742.

Huang, F., Chi, Y., Meng, Q., Gai, J., and Yu, D. 2006. 'GmZFP1 encoding a single zinc finger protein is expressed with enhancement in reproductive organs and late seed development in soybean (*Glycine max*).' *Molecular Biology Reports* 33(4): 279-285.

Hughes, C. E., and Atchison, G. W. 2015. 'The ubiquity of alpine plant radiations: from the Andes to the Hengduan mountains.' *New Phytologist* 207(2): 275–82.

Hugouvieux, V., Silva, C. S., Jourdain, A., Stigliani, A., Charras, Q., Conn, V., Conn, S. J., Carles, C. C., Parcy, F., and Zubieta, C. 2019. 'Tetramerization of MADS family transcription factors SEPALLATA3 and AGAMOUS Is required for floral meristem determinacy in *Arabidopsis*.' *Nucleic Acids Research* 46(10): 4966–77.

Huijser, P., Klein, J., Lönnig, W. E., Meijer, H., Saedler, H., and Sommer, H. 1992. 'Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene squamosa in *Antirrhinum majus*.' *EMBO Journal* 11(4): 1239–49.

Hussein, G. M., Abu El-Heba, G. A., Abdou, S. M., and Abdallah, N. A. 2013. 'Optimization of transient gene expression system in *Gerbera jemosonii* petals.' *GM crops and food* 4(1): 50–57.

Hussey, S. G., Mizrachi, E., Creux, N. M., and Myburg, A. A. 2013. 'Navigating the transcriptional roadmap regulating plant secondary cell wall deposition.' *Frontiers in Plant Science* 4: 325.

Ihlenfeldt, H. D. 1994. 'Diversification in an arid world: the Mesembryanthemaceae.' *Annual Review of Ecology and Systematics* 25: 521–46.

Inagaki, Y., Johzuka-Hisatomi, Y., Mori, T., Takahashi, S., Hayakawa, Y., Peyachoknagul, S., Ozeki, Y., and Iida, S. 1999. 'Genomic organization of the genes encoding dihydroflavonol 4-reductase for flower pigmentation in the Japanese and Common Morning Glories.' *Gene* 226(2): 181–88.

Inkscape. 2020. 'Inkscape Project.' https://inkscape.org.

Irish, V. F. 2008. 'The *Arabidopsis* petal: A model for plant organogenesis.' *Trends in Plant Science* 13(8): 430–36.

Irish, V. F. 2009. 'Evolution of petal identity.' *Journal of Experimental Botany* 60(9): 2517-2527.

Irish, V. F. 2010. 'The flowering of *Arabidopsis* flower development.' *The Plant Journal* 61(6): 1014–28.

Irish, V. F., and Sussex, I. M. 1990. 'Function of the *Apetala-1* gene during *Arabidopsis* floral development.' *The Plant Cell* 2(8): 741–53.

Jacob, F. 1977. 'Evolution and tinkering.' *Science* 196(4295): 1161–66.

de Jager, M. L., and Ellis. A. G. 2017. 'Evolutionary history of a keystone pollinator parallels the biome occupancy of angiosperms in the Greater Cape Floristic Region.' *Molecular Phylogenetics and Evolution* 107: 530–37.

de Jager, M. L., and Ellis, A. G. 2014. 'Floral polymorphism and the fitness implications of attracting pollinating and florivorous insects.' *Annals of Botany* 113(2): 213–22.

de Jager, M. L., and Ellis, A. G. 2012. 'Gender-specific pollinator preference for floral traits.' *Functional Ecology* 26(5): 1197–1204.

de Jager, M. L., and Ellis, A. G. 2013. 'The influence of pollinator phylogeography and mate preference on floral divergence in a sexually deceptive daisy.' *Evolution* 67(6): 1706–14.

Jain, M., and Khurana, J. P. 2009. 'Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice.' *The FEBS Journal* 276(11): 3148-3162.

James, P., Halladay, J., and Craig, E. A. 1996. 'Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast.' *Genetics* 144(4): 1425–36.

Jeffrey, C. 2009. 'Evolution of Compositae flowers.' In *Systematics, Evolution, and Biogeography of Compositae*, Vienna: International Association for Plant Taxonomy, 131–38.

Jia, L., Clegg, M. T., and Jiang, T. 2003. 'Excess non-synonymous substitutions suggest that positive selection episodes occurred during the evolution of DNA-binding domains in the *Arabidopsis* R2R3-MYB gene family.' *Plant Molecular Biology* 52(3): 627–42.

Jiang, C. K., and Rao, G. 2020. 'Insights into the diversification and evolution of R2R3-MYB transcription factors in plants.' *Plant Physiology* 183(2): 637–55.

Jiang, C., Gu, X., and Peterson, T. 2004. 'Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*.' *Genome Biology* 5(7): 1–11.

Jiang, W. K., Liu, Y. L., Xia, E. H., and Gao, L. Z. 2013. 'Prevalent role of gene features in determining

evolutionary fates of whole-genome duplication duplicated genes in flowering plants.' *Plant Physiology* 161(4): 1844–61.

Jiao, Y., and Paterson, A. H. 2014. 'Polyploidy-associated genome modifications during land plant evolution.' *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1648): 20130355.

Jiggins, C. D., Wallbank, R., and Hanly, J. J. 2016. 'Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns?' *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1713).

Johnson, E. T., Ryu, S., Yi, H., Shin, B., Cheong, H., and Choi, G. 2001. 'Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase.' *The Plant Journal* 25(3): 325–33.

Johnson, S. D., and Midgley, J. J. 1997. 'Fly pollination of *Gorteria diffusa* (Asteraceae), and a possible mimetic function for dark spots on the capitulum.' *American Journal of Botany* 84(4): 429–36.

Johnson, S. D. 2010. 'The pollination niche and its role in the diversification and maintenance of the Southern African flora.' *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1539): 499–516.

Johzuka-Hisatomi, Y., Hoshino, A., Mori, T., Habu, Y., and Iida, S. 1999. 'Characterization of the chalcone synthase genes expressed in flowers of the common and Japanese Morning Glories.' *Genes and Genetic Systems* 74(4): 141–47.

Kaczorowski, R. L., Seliger, A. R., Gaskett, A. C., Wigsten, S. K., and Raguso, R. A. 2012. 'Corolla shape vs. size in flower choice by a octurnal Hahwkmoth pollinator.' *Functional Ecology* 26(3): 577–87.

Kang, H. G., Foley, R. C., Oñate-Sánchez, L., Lin, C., and Singh, K. B. 2003. 'Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid.' *The Plant Journal* 35(3): 362-372.

Karis, P. O. 2007. 'Tribe Arctotideae Cass.' In *The Families and Genera of Vascular Plants, Vol. VIII*, Berlin: Springer-Verlag, 223–29.

Karis, P. O. 2009. 'Arctotideae.' In *Funk, V.A., Susanna, A., Stuessy, T.F. and Bayer, R.J. (Eds). Systematics, Evolution and Biogeography Of Compositae.*, Vienna: International Association for Plant Taxonomy, 387–410.

Karis, P. O. 2006. 'Morphological data indicates two major clades of the subtribe Gorteriinae (Asteraceae-Arctotideae).' *Cladistics* 22(3): 199–221.

Kassambara, A., and Mundt, F. 2017. Factoextra: extract and visualize the results of multivariate data analyses. *R package version*, 1(5):337-354.

Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. 2012. 'Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*.' *BMC Genomics* 13(1): 544.

Kato, K., Ohta, K., Komata, Y., Araki, T., Kanahama, K., and Kanayama, Y. 2005. 'Morphological and molecular analyses of the tomato floral mutant leafy inflorescence, a new allele of falsiflora.' *Plant Science* 169(1): 131–38.

Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. 2002. 'MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform.' *Nucleic Acids Research* 30(14): 3059–66.

Kay, K. M., and Sargent, R. D. 2009. 'The role of animal pollination in plant speciation: integrating

ecology, geography, and genetics.' *Annual Review of Ecology, Evolution, and Systematics* 40(1): 637–56.

Kelber, A., Vorobyev, M., and Osorio, D. 2003. 'Animal colour vision – behavioural tests and physiological concepts.' *Biological Reviews* 78: 81–118.

Kelemen, Z., Sebastian, A., Xu, W., Grain, D., Salsac, F., Avon, A., Berger, N., Tran, J., Dubreucq, B., Lurin, C., and Lepiniec, L. 2015. 'Analysis of the DNA-binding activities of the *Arabidopsis* R2R3-MYB transcription factor family by one-hybrid experiments in yeast.' *PLoS ONE* 10(10): 1–22.

Keller, I., and Seehausen, O. 2012. 'Thermal adaptation and ecological speciation.' *Molecular Ecology* 21(4): 782–99.

Kemp, J. E., Bergh, N.G., Soares, M., and Ellis, A. G. 2019. 'Dominant pollinators drive non-random community assembly and shared flower colour patterns in daisy communities.' *Annals of Botany* 123(2): 277–88.

Kersting, A. R., Bornberg-Bauer, E., Moore, A. D., and Grath, S. 2012. 'Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution.' *Genome Biology and Evolution* 4(3): 316–29.

Kim, J., Yi, H., Choi, G., Shin, B., Song, P.S., and Choi, G. 2003. 'Functional characterization of phytochrome interacting factor 3 in phytochrome-mediated light signal transduction.' *The Plant Cell* 15(10): 2399-2407.

Kim, S., Soltis, P. S., Wall, K., and Soltis, D. E. 2006. 'Phylogeny and domain evolution in the *APETALA2-like* gene family.' *Molecular Biology and Evolution* 23(1): 107–20.

Kim, M., Cui, M.L., Cubas, P., Gillies, A., Lee, K., Chapman, M. A., Abbott, R. J., and Coen, E. 2008. Regulatory genes control a key morphological and ecological trait transferred between species. *Science*, *322*(5904):1116-1119.

Kim, S. G., Lee, S., Kim, Y. S., Yun, D. J., Woo, J. C., and Park, C. M. 2010. 'Activation tagging of an Arabidopsis SHI-RELATED SEQUENCE gene produces abnormal anther dehiscence and floral development.' *Plant Molecular Biology* 74(4-5): 337-351.

Kim, H. J., Nam, H. G., and Lim, P. O. 2016. 'Regulatory network of NAC transcription factors in leaf senescence.' *Current Opinion in Plant Biology* 33: 48-56.

Kisel, Y., and Barraclough, T. G. 2010. 'Speciation has a spatial scale that depends on levels of gene flow.' *American Naturalist* 175(3): 316–34.

Klahre, U., Gurba, A., Hermann, K., Saxenhofer, M., Bossolini, E., Guerin, P., and Kuhlemeier, C. 2011. 'Pollinator choice in *Petunia* depends on two major genetic loci for floral scent production.' *Current Biology* 21(9): 730–39.

Klak, C., Reeves, G., and Hedderson, T. 2004. 'Unmatched tempo of evolution in southern African semi-desert ice plants.' *Nature* 427(6969): 63–65.

Kobayashi, S., Ishimaru, M., Hiraoka, K., and Honda, C. 2002. 'Myb-related genes of the kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis.' *Planta* 215(6): 924–33.

Koes, R. E., Spelt, C. E., and Mol, J. N. M. 1989. 'The chalcone synthase multigene family of *Petunia hybrida*: differential, light-regulated expression during flower development and UV light induction.' *Plant Molecular Biology* 12(2): 213–25.

Kohli, A., Gahakwa, D., Vain, P., Laurie, D. A., and Christou, P. 1999. 'Transgene expression in rice

engineered through particle bombardment: molecular factors controlling stable expression and transgene silencing.' *Planta* 208(1): 88–97.

Kondo, T., Yamashiki, J., Kawahori, K., and Goto, T. 1989. 'Structure of lobelinin A and B, novel anthocyanins acylated with three and four different organic acids, respectively.' *Tetrahedron Letters* 30(44): 6055–58.

König, P., Giraldo, R., Chapman, L., and Rhodes, D. 1996. 'The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA.' *Cell* 85(1): 125–36.

Kornberg, R. D. 2007. 'The molecular basis of eukaryotic transcription.' *Proceedings of the National Academy of Sciences* 104(32): 12955-12961.

Koseki, M., Goto, K., Masuta, C., and Kanazawa, A. 2005. 'The star-type color pattern in *Petunia hybrida* 'red star' flowers is induced by sequence-specific degradation of chalcone synthase RNA.' *Plant and Cell Physiology* 46(11): 1879–83.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A., 2019. 'RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.' *Bioinformatics* 35(21): 4453–55.

Krejčíková, J., Sudová, R., Lučanová, M., Trávníček, P., Urfus, T., Vít, P., Weiss-Schneeweiss, H., Kolano, B., Oberlander, K., Dreyer, L.L., and Suda, J. 2013. 'High ploidy diversity and distinct patterns of cytotype distribution in a widespread species of *Oxalis* in the Greater Cape Floristic Region.' *Annals of Botany* 111(4): 641–49.

Krizek, B.A., and Fletcher, J. C. 2005. 'Molecular mechanisms of flower development: an armchair guide.' *Nature Reviews Genetics* 6(9): 688–98.

Kulbaba, M. W., Clocher, I. C., and Harder, L. D. 2017. 'Inflorescence characteristics as function-valued traits: analysis of heritability and selection on architectural effects.' *Journal of Systematics and Evolution* 55(6): 559–65.

Labandeira, C. C. 2010. 'The pollination of mid Mesozoic seed plants and the early history of long-proboscid insects.' *Annals of the Missouri Botanical Garden* 97(4): 469–513.

Labandeira, C. C., Kvaček, J., and Mostovski, M. B. 2007. 'Pollination drops, pollen, and insect pollination of Mesozoic gymnosperms.' *Taxon* 56(3): 663–95.

Lai, Y. S., Shimoyamada, Y., Nakayama, M., and Yamagishi, M. 2012. 'Pigment accumulation and transcription of LhMYB12 and anthocyanin biosynthesis genes during flower development in the Asiatic hybrid Lily (*Lilium* Spp.).' *Plant Science* 193–194: 136–47.

Laitinen, R. A., Ainasoja, M., Broholm, S. K., Teeri, T. H., and Elomaa, P., 2008. 'Identification of target genes for a MYB-type anthocyanin regulator in *Gerbera hybrida*.' *Journal of Experimental Botany* 59(13): 3691–3703.

Lane, M A. 1996. 'Pollination biology of Compositae. In: Caligari P. D. S, Hind D. J. N. Eds. Compositae: Biology and Utilization.' In *Aligari PDS, Hind DJN. Eds.*, Kew: Royal Botanical Gardens, 61–80.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. 2017. 'Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses.' *Molecular Biology and Evolution* 34(3): 772–73.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riaño-Pachón, D. M., Corrêa, L. G., Reski, R., Mueller-Roeber, B. and Rensing, S. A. 2010. 'Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: A timeline of loss, gain, expansion, and correlation with

complexity.' *Genome Biology and Evolution* 2(1): 488–503.

Larter, M., Dunbar-Wallis, A., Berardi, A. E., and Smith, S. D. 2018. 'Convergent evolution at the pathway level: predictable regulatory changes during flower color transitions.' *Molecular Biology and Evolution* 35(9): 2159–69.

Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. 2012. 'Inference of population structure using dense haplotype data.' *PLoS Genetics* 8(1): 11–17.

Le Hir, R., and Bellini, C. 2013. 'The plant-specific Dof transcription factors family: new players involved in vascular system development and functioning in *Arabidopsis*.' *Frontiers in Plant Science* 4: 164.

Lehti-Shiu, M. D., Panchy, N., Wang, P., Uygun, S., and Shiu, S. H. 2017. 'Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families.' *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1860(1): 3–20.

Leonard, A. S., Dornhaus, A., and Papaj, D. 2011. 'Forget-me-not: complex floral displays , inter-signal interactions , and pollinator cognition.' *Current Zoology* 57(2): 215–24.

Lewontin, R . C ., and Ken-ichi, K. 1960. 'The evolutionary dynamics of complex polymorphisms.' *Society for the Study of Evolution Stable* 14(4): 458–72.

Lexer, C., Wüest, R. O., Mangili, S., Heuertz, M., Stölting, K. N., Pearman, P. B., Forest, F., Salamin, N., Zimmermann, N. E., and Bossolini, E., 2014. 'Genomics of the divergence continuum in an African plant biodiversity hotspot , I : drivers of population divergence in *Restio capensis* (Restionaceae).' *Molecular Ecology* (23): 4373–86.

Li, Q., Wang, J., Sun, H. Y., and Shang, X. 2014. 'Flower Color Patterning in Pansy (*Viola*×*wittrockiana* Gams.) Is caused by the differential expression of three genes from the anthocyanin pathway in acyanic and cyanic flower areas.' *Plant Physiology and Biochemistry* 84: 134–41. http://dx.doi.org/10.1016/j.plaphy.2014.09.012.

Li, S., and Zachgo, S. 2013. 'TCP3 interacts with R2R3-MYB proteins, promotes flavonoid biosynthesis and negatively regulates the auxin response in *Arabidopsis thaliana*.' *The Plant Journal* 76(6): 901–13.

Li, X. Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C.L.L., and Chu, H.C. 2008. 'Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm.' *PLoS Biology* 6(2): 0365–88.

Licausi, F., Ohme-Takagi, M., and Perata, P. 2013. 'APETALA 2/Ethylene Responsive Factor (AP 2/ERF) transcription factors: mediators of stress responses and developmental programs.' *New Phytologist*, *199*(3): 639-649.

Lim, S. H., Song, J. H., Kim, D. H., Kim, J. K., Lee, J. Y., Kim, Y. M. and Ha, S. H. 2016. 'Activation of anthocyanin biosynthesis by expression of the radish R2R3-MYB transcription factor gene *RsMYB1*.' *Plant Cell Reports* 35(3): 641–53.

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T.K., Espley, R.V., Hellens, R. P., and Allan, A. C. 2010. 'An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae.' *BMC Plant Biology* 10(1): 50.

Linck, E., and Battey, C. J. 2019. 'Minor allele frequency thresholds strongly affect population structure inference with genomic data sets.' *Molecular Ecology Resources* 19(3): 639–47.

Linder, H. P. 2003. 'The radiation of the Cape flora, Southern Africa.' *Biological Reviews* 78(4): 597–638.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. 2008. 'Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*.' *Cell* 133(3): 523–36.

Liu, X., Dinh, T. T., Li, D., Shi, B., Li, Y., Cao, X., Guo, L., Pan, Y., Jiao, Y., and Chen, X., 2014. 'AUXIN RESPONSE FACTOR 3 integrates the functions of AGAMOUS and APETALA 2 in floral meristem determinacy.' *The Plant Journal* 80(4): 629-641.

Liu, C., Thong, Z., and Yu, H. 2009. 'Coming into bloom: the specification of floral meristems.' *Development* 136(20): 3379–91.

Liu, M., Sun, W., Ma, Z., Zheng, T., Huang, L., Wu, Q., Zhao, G., Tang, Z., Bu, T., Li, C., and Chen, H. 2019. 'Genome-wide investigation of the AP2/ERF gene family in tartary buckwheat *(Fagopyum tataricum*).' *BMC Plant Biology* 19(1): 1–19.

Liu, Q., Wang, Z., Xu, X., Zhang, H., and Li, C. 2015b. 'Genome-wide analysis of C2H2 zinc-finger family transcription factors and their responses to abiotic stresses in poplar (*Populus trichocarpa*).' *PLoS ONE* 10(8): 1–25.

Liu, S., Liu, L., Tang, Y., Xiong, S., Long, J., Liu, Z., and Tian, N., 2017. 'Comparative transcriptomic analysis of key genes involved in flavonoid biosynthetic pathway and identification of a flavonol synthase from *Artemisia annua* L.' *Plant Biology* 19(4): 618–29.

Liu, X.F., Xiang, L.L., Yin, X.R., Grierson, D., Li, F., and Chen, K.S. 2015a. 'The identification of a MYB transcription factor controlling anthocyanin biosynthesis regulation in *Chrysanthemum* flowers.' *Scientia Horticulturae* 194: 278–85. http://dx.doi.org/10.1016/j.scienta.2015.08.018.

Liu, X., Dinh, T.T., Li, D., Shi, B., Li, Y., Cao, X., Guo, L., Pan, Y., Jiao, Y. and Chen, X. 2014. 'AUXIN RESPONSE FACTOR 3 integrates the functions of AGAMOUS and APETALA 2 in floral meristem determinacy.' *The Plant Journal* 80(4): 629-641.

Liu, Y., Song, Q., Li, D., Yang, X., and Li, D. 2017b. 'Multifunctional roles of plant dehydrins in response to environmental stresses.' *Frontiers in Plant Science* 8: 1018.

Liu, Y., Wang, X., Mo, T., Yan, Y., Song, Y., Zhao, Y., Li, J., Shi, S., Liu, X., and Tu, P. 2017. 'Identification and functional application of a new malonyltransferase NbMaT1 towards diverse aromatic glycosides from *Nicotiana benthamiana*.' *RSC Advances* 7(34): 21028–35.

Long, M. C., Nagegowda, D. A., Kaminaga, Y., Ho, K. K., Kish, C. M., Schnepp, J., Sherman, D., Weiner, H., Rhodes, D., and Dudareva, N. 2009. 'Involvement of Snapdragon Benzaldehyde Dehydrogenase in Benzoic Acid Biosynthesis.' *The Plant Journal* 59(2): 256–65.

López-González, L., Mouriz, A., Narro-Diego, L., Bustos, R., Martínez-Zapater, J. M., Jarillo, J. A., and Piñeiro, M. 2014. Chromatin-dependent repression of the *Arabidopsis* floral integrator genes involves plant specific PHD-containing proteins. *The Plant Cell* 26(10): 3922-3938.

Lowry, D. B., Cotton Rockwood, R., and Willis, J. H. 2008. 'Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*.' *Evolution* 62(9): 2196–2214.

Ludwig, S. R., Habera, L. F., Dellaporta, S. L., and Wessler, S. R. 1989. 'Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a rotein similar to transcriptional activators and contains the Myc-homology region.' *Proceedings of the National Academy of Sciences of the United States of America* 86(18): 7092–96.

Lundquist, R. C., Close, T. J., and Kado, C. I. 1984. 'Genetic complementation of A*grobacterium tumefaciens* Ti plasmid mutants in the virulence region.' *MGG Molecular and General Genetics* 193(1): 1–7.

Luo, D., Carpenter, R., Vincent, C., Copsey, L., and Coen, E. 1995. 'Origin of floral asymmetry in *Anthirrhinum*.' *Nature* 383: 794–99.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., and Hernández-Hernández, T. 2015. 'A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity.' *New Phytologist* 207(2): 437–53.

Malinsky, M., Trucchi, E., Lawson, D. J., and Falush, D. 2018. 'RADpainter and FineRADstructure: population inference from RADseq data.' *Molecular Biology and Evolution* 35(5): 1284–90.

Mancinelli, A. L. 2020. 'Interaction between light quality and light quantity in the photoregulation of anthocyanin production.' *Plant Physiology* (5291): 1191–95.

Mantel, N. 1967. 'The detection of disease clustering and a generalized regression approach.' *Cancer Research* 27(2): 209–20.

Marazzi, B., and Sanderson, M. J. 2010. 'Large-scale patterns of diversification in the widespread legume genus *Senna* and the evolutionary role of extrafloral nectaries.' *Evolution* 64(12): 3570–92.

Martin, C., Prescott, A., Mackay, S., Bartlett, J., and Vrijlandt, E. 1991. 'Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus.*' *The Plant Journal* 1(1): 37–49.

Martin, C., and Paz-Ares, J. 1997. 'MYB transcription factors in plants.' *Plant Molecular Biology* 13(2): 67–73.

Martín-Trillo, M., and Cubas, P. 2010. 'TCP genes: a family snapshot ten years later.' *Trends in Plant Science* 15(1): 31-39.

Martins, T. R., Berg, J. J., Blinka, S., Rausher, M. D., and Baum, D. A. 2013. 'Precise spatio-temporal regulation of the anthocyanin biosynthetic pathway leads to petal spot formation in *Clarkia gracilis* (Onagraceae).' *New Phytologist* 197(3): 958–69.

Martins, T. R., Jiang, P., and Rausher, M. D. 2017. 'How petals change their spots: cis -regulatory re-wiring in *Clarkia* (Onagraceae).' *New Phytologist* 216(2): 510–18.

Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., and Takahashi, H. 2006. '*Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism.' *The Plant Cell* 18(11): 3235-3251.

Matus, J. T., Cavallini, E., Loyola, R., Höll, J., Finezzo, L., Dal Santo, S., Vialet, S., Commisso, M., Roman, F., Schubert, A., and Alcalde, J. A., 2017. 'A group of grapevine MYBA transcription factors located in chromosome 14 control anthocyanin synthesis in vegetative organs with different specificities compared with the berry color locus.' *The Plant Journal* 91(2): 220–36.

Meier, J. 2019. 'Speciation genomics PlotADMIXTURE.'

Meinhardt, H., and Gierer, A. 2000. 'Pattern formation by local self-activation and lateral inhibition.' *BioEssays* 22(8): 753–60.

Mellers, G. 2016. 'The evolution of morphological diversity in *Gorteria diffusa*.' University of Cambridge.

Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., Schaefer, D. G. and Dolan, L. 2007. 'An ancient mechanism controls the development of cells with a rooting function in land plants.' *Science* 316(5830): 1477-1480.

de Mendoza, A., Sebé-Pedrós, A., Šestak, M. S., Matejčić, M., Torruella, G., Domazet-Lošo, T. and Ruiz-

Trillo, I. 2013. 'Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular tineages.' *Proceedings of the National Academy of Sciences of the United States of America* 110(50): pp.E4858-E4866.

Millard, P. S., Kragelund, B. B., and Burow, M. 2019. 'R2R3 MYB transcription factors – functions outside the DNA-binding domain.' *Trends in Plant Science* 24(10): 934–46.

Millard, P.S., Weber, K., Kragelund, B.B., and Burow, M. 2019. 'Specificity of MYB interactions relies on motifs in ordered and disordered contexts.' *Nucleic Acids Research* 47(18): 9592–9608.

Mitsuda, N., and Ohme-Takagi, M. 2009. 'Functional analysis of transcription factors in *Arabidopsis*.' *Plant and Cell Physiology* 50(7): 1232–48.

Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. 2012. 'AP2/ERF family transcription factors in plant abiotic stress responses.' *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819(2): 86–96.

Moeller, D. A. 2005. 'Pollinator community structure and sources of spatial variation in plant-pollinator interactions in *Clarkia xantiana* ssp*. xantiana*.' *Oecologia* 142(1): 28–37.

Molinero-Rosales, N., Jamilena, M., Zurita, S., Gómez, P., Capel, J., and Lozano, R., 1999. 'FALSIFLORA, the tomato orthologue of FLORICAULA and LEAFY, controls flowering time and Floral Meristem Identity.' *The Plant Journal* 20(6): 685–93.

Molitor, A. M., Bu, Z., Yu, Y., and Shen, W. H. 2014. 'Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes.' *PLoS Genet* 10(1): p.e1004091.

Momonoi, K., Yoshida, K., Mano, S., Takahashi, H., Nakamori, C., Shoji, K., Nitta, A., and Nishimura, M. 2009. 'A vacuolar iron transporter in tulip, TgVit1, is responsible for blue coloration in petal cells through iron accumulation.' *The Plant Journal* 59(3): 437–47.

Monniaux, M., and Hay, A. 2016. 'Cells, walls, and endless forms.' *Current Opinion in Plant Biology* 34: 114–21.

Mooney, M., Desnos, T., Harrison, K., Jones, J., Carpenter, R., and Coen, E. 1995. 'Altered regulation of tomato and tobacco pigmentation genes caused by the Delila gene of *Antirrhinum*.' *The Plant Journal* 7(2): 333–39.

Moore, B. R., and Donoghue, M. J. 2007. 'Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations.' *American Naturalist* 170(S2): S28-S55.

Morita, Yasumasa et al. 2006. 'Isolation of cDNAs for R2R3-MYB, bHLH and WDR transcriptional regulators and identification of c and ca mutations conferring white flowers in the Japanese Morning Glory.' *Plant and Cell Physiology* 47(4): 457–70.

Morita, Y., Saito, R., Ban, Y., Tanikawa, N., Kuchitsu, K., Ando, T., Yoshikawa, M., Habu, Y., Ozeki, Y., and Nakayama, M. 2012. 'Tandemly arranged chalcone synthase A genes contribute to the spatially regulated expression of siRNA and the natural bicolor floral phenotype in *Petunia hybrida*.' *The Plant Journal* 70(5): 739–49.

Moyroud, E., Monniaux, M., Thévenon, E., Dumas, R., Scutt, C.P., Frohlich, M.W., and Parcy, F. 2017. 'A link between LEAFY and B-Gene homologues in *Welwitschia mirabilis* sheds light on ancestral mechanisms prefiguring floral development.' *New Phytologist* 216(2): 469–81.

Moyroud, E., and Glover, B. J. 2017. 'The evolution of diverse floral morphologies.' *Current Biology* 27(17): R941–51.

Moyroud, E., Tichtinsky, G., and Parcy, F. 2009. 'The LEAFY floral regulators in angiosperms: conserved proteins with diverse roles.' *Journal of Plant Biology* 52(3): 177–85.

Muchhala, N., and Serrano, D. 2015. 'The complexity of background clutter affects nectar bat use of flower odor and shape cues.' *PLoS ONE* 10(10): 1–12.

Mukhopadhyay, A., Vij, S., and Tyagi, A. K. 2004. 'Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco.' *Proceedings of the National Academy of Sciences of the United States of America* 101(16): 6309-6314.

Musker, S. D., Ellis, A. G., Schlebusch, S. A., and Verboom, G. A. 2020. 'Niche specificity influences gene flow across fine-scale habitat mosaics in Succulent Karoo plants.' *Molecular Ecology* 30(1): 175–92.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., and Kent, J. 2000. 'Biodiversity hotspots for conservation priorities.' *Nature* 403(6772): 853–58.

Nadeau, J.A. 2009. 'Stomatal development: new signals and fate determinants.' *Current Opinion in Plant Biology* 12(1): 29-35.

Nagaraju, V., Srinivas, G. S. L., and Lakshmi Sita, G. 1998. '*Agrobacterium*-mediated genetic transformation in *Gerbera hybrida*.' *Current Science* 74(7): 630–34.

Nakai, Y., Nakahira, Y., Sumida, H., Takebayashi, K., Nagasawa, Y., Yamasaki, K., Akiyama, M., Ohme-Takagi, M., Fujiwara, S., Shiina, T., and Mitsuda, N. 2013. 'Vascular plant one-zinc-finger protein 1/2 transcription factors regulate abiotic and biotic stress responses in *Arabidopsis*.' *The Plant Journal* 73(5): 761–75.

Nakano, T., Suzuki, K., Ohtsuki, N., Tsujimoto, Y., Fujimura, T., and Shinshi, H. 2006. 'Identification of genes of the plant-specific transcription-factor families cooperatively regulated by ethylene and jasmonate in *Arabidopsis thaliana*.' *Journal of Plant Research* 119(4): 407-413.

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K., 2012. 'NAC transcription factors in plant abiotic stress responses.' *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819(2): 97–103.

Nakatsuka, A., Izumi, Y., and Yamagishi, M., 2003. Spatial and temporal expression of *chalcone synthase* and *dihydroflavonol 4-reductase* genes in the Asiatic hybrid lily. *Plant Science*, 165(4): 759-767.

Nakatsuka, T., Yamada, E., Saito, M., Fujita, K., and Nishihara, M., 2013. 'Heterologous expression of *Gentian MYB1R* transcription factors suppresses anthocyanin pigmentation in tobacco flowers.' *Plant Cell Reports* 32(12): 1925–37.

Nakayama, T., Suzuki, H., and Nishino, T. 2003. 'Anthocyanin acyltransferases: specificities, mechanism, phylogenetics, and applications.' *Journal of Molecular Catalysis B: Enzymatic* 23(2–6): 117–32.

Narbona, E., Jaca, J., Del Valle, J. C., Valladares, F., and Buide, M. L. 2018. 'Whole-plant reddening in *Silene germana* is due to anthocyanin accumulation in response to visible light.' *Plant Biology* 20(6): 968–77.

Nathan, R., and Muller-landau, H. C. 2000. 'Spatial patterns of seed dispersal.' *Trends in Ecology and Evolution* 15(7): 278–85.

Navarro, A., and Barton, N. H. 2003. 'Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation.' *Evolution* 57(3): 447–59.

Van der Graaff, E., Laux, T., and Rensing., S. A. 2009. 'Protein family review the WUS Homeobox-containing (WOX) protein family.' *Genome Biology* 10(12): 1–9.

Van der Niet, T., Pirie, M.D., Shuttleworth, A., Johnson, S.D., and Midgley, J.J., 2014. 'Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape Shrub *Erica plukenetii*?' *Annals of Botany* 113(2): 301–15.

Van der Niet, T., and Johnson, S. D. 2012. 'Phylogenetic evidence for pollinator-driven diversification of angiosperms.' *Trends in Ecology and Evolution* 27(6): 353–61.

Van der Niet, T., Peakall, R., and Johnson, S. D. 2014. 'Pollinator-driven ecological speciation in plants: new evidence and future perspectives.' *Annals of Botany* 113(2): 199–211.

Noda, K., Glover, B. J., Linstead, P., and Martin, C. 1994. 'Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor.' *Nature* 369(6482): 661–64.

Nørbæk, R., and Kondo, T. 1999. 'Anthocyanins from flowers of *Lilium* (Liliaceae).' *Phytochemistry* 50(7): 1181–84.

Nosil, P., Vines, T. H., and Funk, D. J. 2005. 'Immigrants from divergent habitats.' *Evolution; International Journal of Organic Evolution* 59: 705–19.

Nunes, M. D. S, Arif, S., Schlotterer, C., and McGregor, A. P. 2013. 'A perspective on micro-evo-devo: progress and potential.' *Genetics* 195(3): 625–34.

O'Malley, R. C., Huang, S. S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., and Ecker, J. R. 2016. 'Cistrome and epicistrome features shape the regulatory DNA landscape.' *Cell* 165(5): 1280–92.

O'Meara, B. C., Smith, S. D., Armbruster, W. S., Harder, L. D., Hardy, C. R., Hileman, L. C., Hufford, L., Litt, A., Magallón, S., Smith, S. A., and Stevens, P. F. 2016. 'Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity.' *Proceedings of the Royal Society B: Biological Sciences* 283(1830): p.20152304.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. 1992. 'Solution structure of a DNA-binding unit of Myb: A helix-turn-helix-related motif with conserved Tryptophans forming a hydrophobic core.' *Proceedings of the National Academy of Sciences of the United States of America* 89(14): 6428–32.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. 1994. 'Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices.' *Cell* 79(4): 639–48.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., and Ishii, S. 1995. 'Comparisons of the free and DNA-complexed forms of the DNA-binding domain from c-Myb.' *Nature Structural Biology* 2(4): 309–20.

Oh, E., Kim, J., Park, E., Kim, J. I., Kang, C. and Choi, G. 2004. 'PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in *Arabidopsis thaliana*.' *The Plant Cell* 16(11): 3045-3058.

Ohno, S., Hosokawa, M., Kojima, M., Kitamura, Y., Hoshino, A., Tatsuzawa, F., Doi, M., and Yazawa, S.2011. 'Simultaneous post-transcriptional gene silencing of two different chalcone synthase genes resulting in pure white flowers in the octoploid *Dahlia*.' *Planta* 234(5): 945–58.

Ollerton, J., Winfree, R., and Tarrant, S. 2011. 'How many flowering plants are pollinated by animals?'

Oikos 120(3): 321–26.

Onstein, R. E., Carter, R. J., Xing,Y., and Linder, H. P. 2014. 'Diversification rate shifts in the Cape Floristic Region: the right traits in the right place at the right time.' *Perspectives in Plant Ecology, Evolution and Systematics* 16(6): 331–40.

Orteu, A., and Jiggins, C. D. 2020. 'The genomics of coloration provides insights into adaptive evolution.' *Nature Reviews Genetics*: 1–15.

Osawa, Y. 1982. 'Copigmentation of anthocyanins.' In *In P. Markakis (Ed.) Anthocyanins as Food Colors*, New York Academic Press, 41–69.

Owen, C. R., and Bradshaw, H. D. 2011. 'Induced mutations affecting pollinator choice in *Mimulus lewisii* (Phrymaceae).' *Arthropod-Plant Interactions* 5(3): 235–44.

Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., and Weigel, D., 2003. 'Control of leaf morphogenesis by microRNAs.' *Nature* 425(6955): 257-263.

Pandey, A., Misra, P., Bhambhani, S., Bhatia, C., and Trivedi, P.K. 2014. 'Expression of *Arabidopsis* Myb transcription factor, Atmyb111, in tobacco requires light to modulate flavonol content.' *Scientific Reports* 4(1): 1-10.

Panero, J. L., and Funk, V. A. 2008. 'The value of sampling anomalous taxa in phylogenetic studies : major clades of the Asteraceae revealed.' *Molecular Phylogenetics and Evolution* 47(2): 757–82.

Papiorek, S., Junker, R. R., Alves-dos-Santos, I., Melo, G. A., Amaral-Neto, L. P., Sazima, M., Wolowski, M., Freitas, L., and Lunau, K. 2016. 'Bees, birds and yellow flowers: pollinator-dependent convergent evolution of UV patterns.' *Plant Biology* 18(1): 46–55.

Paris, J. R., Stevens, J. R., and Catchen, J. M. 2017. 'Lost in parameter space : A road map for STACKS.' *Methods in Ecology and Evolution* 8(10): 1360–73.

Parolin, P. 2001. 'Seed expulsion in fruits of Mesembryanthema (Aizoaceae): A mechanistic approach to study the effect of fruit morphological structures on seed dispersal.' *Flora* 196(4): 313–22.

Parolin, P. 2006. 'Ombrohydrochory: rain-operated seed dispersal in plants - with special regard to jet-action dispersal in Aizoaceae.' *Flora: Morphology, Distribution, Functional Ecology of Plants* 201(7): 511–18.

Pattanaik, S., Kong, Q., Zaitlin, D., Werkman, J. R., Xie, C. H., Patra, B., and Yuan, L. 2010. 'Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco.' *Planta* 231(5): 1061–76.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A., and Saedler, H. 1987. 'The regulatory C1 locus of *Zea mays* encodes a protein with homology to Myb proto-oncogene products and with structural similarities to transcriptional activators.' *The EMBO Journal* 6(12): 3553–58.

Paz-Ares, J., Wienand, U., Peterson, P. A., and Saedler, H. 1986. 'Molecular cloning of the c locus of *Zea mays*: A locus regulating the anthocyanin pathway.' *The EMBO Journal* 5(5): 829–33.

Peccoud, J., and Simon, J. 2010. 'The pea aphid complex as a model of ecological speciation.' *Ecological Entomology* 35: 119–30.

Peitsch, D., Fietz, A., Hertel, H., de Souza, J., Ventura, D. F., and Menzel, R. 1992. 'The spectral input systems of hymenopteran insects and their receptor-based colour vision.' *Journal of Comparative Physiology A* 170: 23–40.

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. 2000. 'B and C floral organ identity

functions require SEPALLATA MADS-box genes.' *Nature* 405(6783): 200–203.

Pelletier, M. K., and Winkel-Shirley, B. 1996. 'Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings.' *Plant Physiology* 111(1): 339–345.

Pérez-Barrales, R., Arroyo, J., and Armbruster, W. S. 2007. 'Differences in pollinator faunas may generate geographic differences in floral morphology and integration in *Narcissus papyraceus* (Amaryllidaceae).' *Oikos* 116(11): 1904–18.

Pesch, M., Schultheiß, I., Klopffleisch, K., Uhrig, J. F., Koegl, M., Clemen, C. S., Simon, R., Weidtkamp-Peters, S., and Hülskamp, M. 2015. 'Transparent testa Glabra1 and Glabra1 compete for binding to Glabra3 in *Arabidopsis*.' *Plant Physiology* 168(2): 584–97.

Peter, B. M., and Slatkin, M. 2013. 'Detecting range expansions from genetic data.' *Evolution* 67(11): 3274–89.

Petit, P., Granier, T., d'Estaintot, B. L., Manigand, C., Bathany, K., Schmitter, J. M., Lauvergeat, V., Hamdi, S. and Gallois, B. 2007. 'Crystal structure of grape dihydroflavonol 4-reductase , a key enzyme in flavonoid biosynthesis.' *Journal of Molecular Biology* 368: 1345–57.

Petroni, K., and Tonelli, C. 2011. 'Recent advances on the regulation of anthocyanin synthesis in reproductive organs.' *Plant Science* 181(3): 219–29.

Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., and Vianello, A. 2013. 'Plant flavonoids-biosynthesis, transport and involvement in stress responses.' *International Journal of Molecular Sciences* 14(7): 14950–73.

Pfaffl, M. W. 2004. 'Quantification strategies in real-time PCR.' In *A-Z of Quantitative PCR*, 89–113.

Phukan, U. J., Jeena, G. S., Tripathi, V., and Shukla, R. K. 2017. 'Regulation of Apetala2/Ethylene response factors in plants.' *Frontiers in Plant Science* 8: 150.

Pimm, S. L., and Raven, P. H. 2017. 'The fate of the world's plants.' *Trends in Ecology and Evolution* 32(5): 317–20.

Pinheiro, J., Bates, D., DebRoy, S., and Deepayan, S. 2012. 'nlme: Linear and nonlinear mixed effects models.' *R package version 3.1-108*.

Posé, D., Yant, L., and Schmid, M. 2012. 'The end of innocence: flowering networks explode in complexity.' *Current Opinion in Plant Biology* 15(1): 45–50.

Preston, J. C., Barnett, L. L., Kost, M. A., Oborny, N. J., and Hileman, L. C. 2014. 'Optimization of virus-induced gene silencing to facilitate evo-devo studies in the emerging model species *Mimulus guttatus* (Phrymaceae).' *Annals of the Missouri Botanical Garden* 99(3): 301–12.

Prouse, M. B., and Campbell, M. M. 2012. 'The interaction between MYB proteins and their target DNA binding sites.' *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819(1): 67–77.

Prud'homme, B., Gompel, N., and Carroll, S. B. 2007. 'Emerging principles of regulatory evolution.' *Proceedings of the National Academy of Sciences of the United States of America* 104: 8605–12.

Prunier, R., Akman, M., Kremer, C. T., Aitken, N., Chuah, A., Borevitz, J., and Holsinger, K. E. 2017. 'Isolation by distance and isolation by environment contribute to population differentiation in *Protea repens* (Proteaceae L.), a widespread South African species.' *American Journal of Botany* 104(5): 674–84.

Pu, X., and Goodman, R. N. 1992. 'Induction of necrogenesis by *Agrobacterium tumefaciens* on grape

explants.' *Physiological and Molecular Plant Pathology* 41(4): 241–54.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., and Sham, P. C. 2007. 'PLINK: A tool set for whole-genome association and population-based linkage analyses.' *American Journal of Human Genetics* 81(3): 559–75.

Purcell, S., and Chang, C. 2017. 'PLINK 1.9.' http://pngu.mgh.harvard.edu/purcell/plink/.

Qi, T., Song, S., Ren, Q., Wu, D., Huang, H., Chen, Y., Fan, M., Peng, W., Ren, C., and Xie, D. 2011. 'The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*.' *The Plant Cell* 23(5): 1795-1814.

Qin, Y., Ma, X., Yu, G., Wang, Q. I., Wang, L., Kong, L., Kim, W., and Wang, H.W. 2014. 'Evolutionary history of trihelix family and their functional diversification.' *DNA Research* 21(5): 499–510.

Quattrocchio, F., Wing, J. F., Leppen, H. T., Mol, J. N., and Koes, R. E. 1993. 'Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes.' *The Plant Cell* 5(11): 1497–1512.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J. and Koes, R. 1999. 'Molecular Analysis of the *Anthocyanin2* Gene of *Petunia* and Its Role in the Evolution of Flower Color.' *The Plant Cell* 11(8): 1433–1444.

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J. and Koes, R. 2006. 'PH4 of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway.' *The Plant Cell* 18(5): 1274–91.

Rafique, M. Z., Carvalho, E., Stracke, R., Palmieri, L., Herrera, L., Feller, A., Malnoy, M. and Martens, S. 2016. 'Nonsense mutation inside anthocyanidin synthase gene controls pigmentation in yellow raspberry (*Rubus Idaeus* L .).' *Frontiers in Plant Science* 7: 1–12.

Raguso, R. A. 2004. 'Flowers as sensory billboards: progress towards an integrated understanding of floral advertisement.' *Current Opinion in Plant Biology* 7(4): 434–40.

Raguso, R. A. 2008. 'Wake up and smell the roses: the ecology and evolution of floral scent.' *Annual Review of Ecology, Evolution, and Systematics* 39(1): 549–69.

Raguso, R. A., and Willis, M. A. 2005. 'Synergy between visual and olfactory cues in nectar feeding by wild hawkmoths , *Manduca sexta*.' *Animal Behaviour* 69(2): 407–18.

Ramsay, N. A., and Glover, B. J. 2005. 'MYB-BHLH-WD40 protein complex and the evolution of cellular diversity.' *Trends in Plant Science* 10(2): 63–70.

Ramsey, J., and Schemske, D. W. 1998. 'Pathways, mechanisms, and rates of polyploid formation in flowering plants.' *Annual Review of Ecology and Systematics* 29(1): 467–501.

Ramsey, J., and Schemske, D. W. 2002. 'Neopolyploidy in flowering plants.' *Annual Review of Ecology and Systematics* 33: 589–639.

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B. and Westram, A. M. 2017. 'Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow.' *Journal of Evolutionary Biology* 30(8): 1450–77.

Rawat, R., Schwartz, J., Jones, M.A., Sairanen, I., Cheng, Y., Andersson, C.R., Zhao, Y., Ljung, K., and Harmer, S.L. 2009. 'REVEILLE 1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways.' *Proceedings of the National Academy of Sciences of the United States of America* 106(39): 16883–88.

Rebeiz, M., Patel, N. H., and Hinman, V. F. 2015. 'Unraveling the tangled skein: the evolution of transcriptional regulatory networks in development.' *Annual Review of Genomics and Human Genetics* 16: 103–31.

Ren, M. X., and Tang. J. Y. 2010. 'Anther fusion enhances pollen removal in *Campsis grandiflora*, a hermaphroditic flower with Didynamous stamens.' *International Journal of Plant Sciences* 171(3): 275–82.

Rensing, S. A. 2014. 'Gene duplication as a driver of plant morphogenetic evolution.' *Current Opinion in Plant Biology* 17(1): 43–48.

Resentini, F., Felipo-Benavent, A., Colombo, L., Blázquez, M.A., Alabadí, D., and Masiero, S. 2015. 'TCP14 and TCP15 mediate the promotion of seed germination by gibberellins in *Arabidopsis thaliana*.' *Molecular Plant* 8(3): 482-485.

Rey, P. J., Herrera, C. M., Guitián, J., Cerdá, X., Sanchez-Lafuente, A. M., Medrano, M., and Garrido, J. L. 2006. 'The geographic mosaic in predispersal interactions and selection on *Helleborus foetidus* (Ranunculaceae).' *Journal of Evolutionary Biology* 19(1): 21–34.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D. O., Truco, M. J., and Xia, R. 2017. 'Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce.' *Nature Communications* 8(1): 1-11.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., and Creelman, R. 2000. '*Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes.' *Science* 290(5499): 2105–10.

Rieseberg, L. H. 2001. 'Chromosomal rearrangements and speciation.' *TRENDS in Ecology and Evolution* 16(7): 351–58.

Rinaldo, A. R., Cavallini, E., Jia, Y., Moss, S. M., McDavid, D. A., Hooper, L. C., Robinson, S. P., Tornielli, G. B., Zenoni, S., Ford, C. M., and Boss, P. K., 2015. 'A grapevine anthocyanin acyltransferase, transcriptionally regulated by VvMYBA, can produce most acylated anthocyanins present in grape skins.' *Plant Physiology* 169(3): 1897–1916.

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. 2019. 'Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics.' *Molecular Ecology* 28(21): 4737–54.

Roessler, H. 1959. *Revision Der Arctotideae-Gorteriinae (Compositae)*. München: Mitteilungen der Botanischen Staatssammlung.

Roessler, H. 1973. *Nachträge Zur Bearbeitung Der Arctotideae-Gorteriinae (Compositae)*. München: Mitteilungen der Botanischen Staatssammlung.

Ronce, O. 2007. 'How Does It Feel to Be like a Rolling Stone? Ten questions about dispersal evolution.' *Annual Review of Ecology, Evolution, and Systematics* 38: 231–53.

Rottmann, W. H., Meilan, R., Sheppard, L. A., Brunner, A. M., Skinner, J. S., Ma, C., Cheng, S., Jouanin, L., Pilate, G. and Strauss, S. H. 2000. 'Diverse effects of overexpression of LEAFY and PTLF, a poplar (*Populus*) homolog of LEAFY/FLORICAULA, in transgenic poplar and *Arabidopsis*.' *The Plant Journal* 22(3): 235–45.

Rowlands, T., Baumann, P., and Jackson, S. P. 1994. 'The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria.' *Science* 264(5163): 1326-1329.

Rueda-Romero, P., Barrero-Sicilia, C., Gómez-Cadenas, A., Carbonero, P., and Oñate-Sánchez, L. 2012. '*Arabidopsis thaliana* DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14.' *Journal of Experimental Botany* 63(5): 1937-1949.

Rundle, H. D., and Nosil, P. 2005. 'Ecological speciation.' *Ecology Letters* 8(3): 336–52.

Sagawa, J. M., Stanley, L. E., LaFountain, A. M., Frank, H. A., Liu, C. and Yuan, Y. W. 2016. 'An R2R3-MYB transcription factor regulates carotenoid pigmentation in *Mimulus lewisii* flowers.' *New Phytologist* 209(3): 1049–57.

Sainsbury, S., Bernecky, C., and Cramer, P. 2015. 'Structural basis of transcription initiation by RNA polymerase II.' *Nature Reviews Molecular Cell Biology* 16(3): 129-143.

Salih, H., He, S., Li, H., Peng, Z., and Du, X. 2020. 'Investigation of the EIL / EIN3 transcription factor gene family members and their expression levels.' *Plants* 9(1): 128.

Saito, R., Fukuta, N., Ohmiya, A., Itoh, Y., Ozeki, Y., Kuchitsu, K., and Nakayama, M. 2006. 'Regulation of anthocyanin biosynthesis involved in the formation of marginal picotee petals in *Petunia*.' *Plant Science* 170(4): 828–34.

Sargent, R. D. 2004. 'Floral symmetry affects speciation rates in angiosperms.' *Proceedings of the Royal Society B: Biological Sciences* 271(1539): 603–8.

Sauquet, H., and Magallón, S. 2018. 'Key questions and challenges in angiosperm macroevolution.' *New Phytologist* 219(4): 1170–87.

Sauret-Güeto, S., Schiessl, K., Bangham, A., Sablowski, R., and Coen, E. 2013. 'JAGGED controls *Arabidopsis* petal growth and shape by interacting with a divergent polarity field.' *PLoS Biology* 11(4) e1001550.

Schemske, D. W. 2009. 'Biotic interactions and speciation in the tropics.' In *Speciation and Patterns of Diversity*, 219–39.

Schemske, D. W., and Bradshaw, H. D. 1999. 'Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*).' *Proceedings of the National Academy of Sciences of the United States of America,* 96(21): 11910-11915.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., and Tinevez, J. Y. 2012. 'Fiji: An open-source platform for biological-image analysis.' *Nature Methods* 9(7): 676–82.

Schluter, D. 2009. 'Evidence for ecological speciation and its alternative.' *Science* 323(5915): 737–41.

Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C. P., Mackay, S. and Talianidis, I. 2010. 'Five-vertebrate ChIP-Seq reveals the evolutionary dynamics of transcription factor binding.' *Science* 328(5981): 1036–40.

Schmidt, G. W., and Delaney, S. K. 2010. 'Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress.' *Molecular Genetics and Genomics* 283(3): 233–41.

Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E. E., Nath, U., and Weigel, D. 2008. 'Control of jasmonate biosynthesis and senescence by miR319 targets.' *PLoS Biol* 6(9): p.e230.

Schurr, F. M., Bond, W. J., Midgley, G. F., and Higgins, S. I. 2005. 'A mechanistic model for secondary seed dispersal by wind and its experimental validation.' *Journal of Ecology* 93(5): 1017–28.

Schütz, K., Persike, M., Carle, R., and Schieber, A. 2006. 'Characterization and quantification of anthocyanins in selected artichoke (*Cynara scolymus* L.) cultivars by HPLC-DAD-ESI-MS.' *Analytical and Bioanalytical Chemistry* 384(7–8): 1511–17.

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K., and Martin, C. 2006. 'A small family of MYB-regulated genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*.' *The Plant Cell* 18: 831–51.

Schwinn, K. E., Ngo, H., Kenel, F., Brummell, D. A., Albert, N. W., McCallum, J. A., Pither-Joyce, M., Crowhurst, R. N., Eady, C., and Davies, K. M. 2016. 'The Onion (*Allium cepa* L.) R2R3-MYB gene *MYB1* regulates anthocyanin biosynthesis.' *Frontiers in Plant Science* 7: 1–16.

Servedio, M. R. 2015. 'Geography, assortative mating, and the effects of sexual selection on speciation with gene flow.' *Evolutionary Applications* 9(1): 91–102.

Sessa, G., Carabelli, M., Possenti, M., Morelli, G., and Ruberti, I. 2018. 'Multiple links between HD-Zip proteins and hormone networks.' *International Journal of Molecular Sciences* 19(12): 4047.

Shaik, Z. 2019. 'Species delimitation and speciation processes in the *Seriphium plumosum* L. complex (Gnaphalieae: Asteraceae) in South Africa.' University of Cape Town, PhD thesis.

Shan, H., Zahn, L., Guindon, S., Wall, P. K., Kong, H., Ma, H., Depamphilis, C. W., and Leebens-Mack, J. 2009. 'Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications.' *Molecular Biology and Evolution* 26(10): 2229–44.

Shan, H., Cheng, J., Zhang, R., Yao, X. and Kong, H. 2019. 'Developmental mechanisms involved in the diversification of flowers.' *Nature Plants* 5(9): 917–23.

Shang, Y., Venail, J., Mackay, S., Bailey, P. C., Schwinn, K. E., Jameson, P. E., Martin, C. R., and Davies, K. M. 2011. 'The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of *Antirrhinum.*' *New Phytologist* 189(2): 602–15.

Sheehan, H., Hermann, K., and Kuhlemeier, C. 2012. 'Color and scent: how single genes influence pollinator attraction.' *Cold Spring Harbor Symposia on Quantitative Biology* 77: 117–33.

Shi, M., and De-Yu, X. 2014. 'Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*.' *Recent Patents on Biotechnology* 8(1): 47–60.

Shimada, N., Sasaki, R., Sato, S., Kaneko, T., Tabata, S., Aoki, T., and Ayabe, S. I., 2005. 'A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the *Lotus japonicus* genome.' *Journal of Experimental Biology* 56(419): 2573–85.

Shiu, S. H., Shih, M. C., and Li, W. H. 2005. 'Transcription factor families have much higher expansion rates in plants than in animals.' *Plant Physiology* 139(1): 18–26.

Shoji, K., Miki, N., Nakajima, N., Momonoi, K., Kato, C., and Yoshida, K., 2007. 'Perianth bottom-specific blue color development in tulip cv. murasakizuisho requires ferric ions.' *Plant and Cell Physiology* 48(2): 243–51.

Shoji, K., Momonoi, K., and Tsuji, T. 2010. 'Alternative expression of vacuolar iron transporter and ferritin genes leads to blue/purple coloration of flowers in tulip Cv. 'murasakizuisho.'' *Plant and Cell Physiology* 51(2): 215–24.

Silvestro, D., Cascales-Miñana, B., Bacon, C. D., and Antonelli, A. 2015. 'Revisiting the origin and diversification of vascular plants through a comprehensive bayesian analysis of the fossil record.' *New Phytologist* 207(2): 425–36.

Singh, K. B., Foley, R. C., and Oñate-Sánchez, L. 2002. 'Transcription factors in plant defense and stress responses.' *Current Opinion in Plant Biology* 5(5): 430–36.

Slatkin, M. 1973. 'Gene flow and sselection in a cline.' *Genetics* 75(4): 733–56.

Slatkin, M. 1985. 'Gene flow in natural populations.' *Annual Review of Ecology and Systematics* 16: 393–430.

Slatkin, M. 1993. 'Isolation by distance in equilibrium and non-equilibrium populations.' *Evolution* 47(1): 264–79.

Smith, C. C., Tittes, S., Mendieta, J. P., Collier-Zans, E., Rowe, H. C., Rieseberg, L. H., and Kane, N. C. 2018. 'Genetics of alternative splicing evolution during sunflower domestication.' *Proceedings of the National Academy of Sciences of the United States of America* 115(26) : 1–6.

Smith, S. D., and Kriebel, R. 2018. 'Convergent evolution of floral shape tied to pollinator shifts in *Iochrominae* (Solanaceae).' *Evolution* 72(3): 6768–6773.

Smith, S. D., and Rausher, M. D. 2011. 'Gene loss and parallel evolution contribute to species difference in flower color.' *Molecular Biology and Evolution* 28(10): 2799–2810.

Snijman, D. A. 2013. 'Plants of the Greater Cape Floristic Region. 2: The extra Cape flora.' In Pretoria: National Biodiversity Institute.

Sobel, J. M., Chen, G. F., Watt, L. R., and Schemske, D. W. 2010. 'The biology of speciation.' *Evolution* 64(2): 295–315.

Sohlberg, J. J., Myrenås, M., Kuusk, S., Lagercrantz, U., Kowalczyk, M., Sandberg, G., and Sundberg, E. 2006. 'STY1 regulates auxin homeostasis and affects apical–basal patterning of the *Arabidopsis* gynoecium.' *The Plant Journal* 47(1): 112-123.

Soltis, P. S., Folk, R. A., and Soltis, D. E. 2019. 'Darwin review: angiosperm phylogeny and evolutionary radiations.' *Proceedings of the Royal Society B: Biological Sciences* 286(1899): 20190099.

Soltis, P. S., and Soltis, D. E. 2004. 'The origin and diversification of angiosperms.' *American Journal of Botany* 91(10): 1614–26.

Song, B. H., and Mitchell-Olds, T. 2011. 'Evolutionary and ecological genomics of non-model plants.' *Journal of Systematics and Evolution* 49(1): 17–24.

Souer, E., van der Krol, A., Kloos, D., Spelt, C., Bliek, M., Mol, J., and Koes, R., 1998. 'Genetic control of branching pattern and floral identity during *Petunia* inflorescence development.' *Development* 125(4): 733–42.

Souer, E., Rebocho, A. B., Bliek, M., Kusters, E., de Bruin, R. A., and Koes, R. 2008. 'Patterning of inflorescences and flowers by the F-Box protein double top and the leafy homolog aberrant leaf and flower of *Petunia*.' *The Plant Cell* 20(8): 2033–48.

Soulebeau, A., Aubriot, X., Gaudeul, M., Rouhan, G., Hennequin, S., Haevermans, T., Dubuisson, J. Y., and Jabbour, F. 2015. 'The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept.' *Organisms Diversity and Evolution* 15(4): 747–61.

Spaethe, J., Tautz, J., and Chittka, L. 2001. 'Visual constraints in foraging bumblebees: flower size and color affect search time and flight behavior.' *Proceedings of the National Academy of Sciences of the United States of America* 98(7): 3898–3903.

Specht, C. D., and Howarth, D. G. 2015. 'Adaptation in flower form: A comparative evodevo approach.' *New Phytologist* 206(1): 74–90.

Spelt, C., Quattrocchio, F., Mol, J. N. M, and Koes, R. 2000. '*Anthocyanin1* of *Petunia* encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes.' *The Plant Cell* 12(9): 1619–31.

Spitzer-Rimon, B., Farhi, M., Albo, B., Cna'ani, A., Zvi, M. M. B., Masci, T., Edelbaum, O., Yu, Y., Shklarman, E., Ovadis, M., and Vainstein, A. 2012. 'The R2R3-MYB-like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating ODO1 and structural scent-related genes in *Petunia*.' *The Plant Cell* 24(12): 5089–5105.

Springob, K., Nakajima, J. I., Yamazaki, M., and Saito, K. 2003. 'Recent advances in the biosynthesis and accumulation of anthocyanins.' *Natural Product Reports* 20(3): 288–303.

Stachel, S. E., and Zambryski, P. C. 1986. 'VirA and VirG control the plant-induced activation of the T-DNA transfer process of *A. tumefaciens*.' *Cell* 46(3): 325–33.

Sigg, C.D., and Buhmann, J.M. 2008. 'Expectation-maximization for sparse and non-negative PCA.' In *Proceedings of the 25th international conference on Machine learning* (pp. 960-967).

Stångberg, F., Ellis, A. G., and Anderberg, A. A. 2013. 'Evolutionary relationships in *Gorteria*: A re-Evaluation.' *Taxon* 62(3): 537–49.

Stångberg, F., and Anderberg, A. A. 2014. 'Morphology and taxonomic reclassification of *Gorteria* (Asteraceae).' *Willdenowia* 44(1):97-120.

Stankowski, S., Sobel, J. M., and Streisfeld, M. A. 2016. 'Geographic cline analysis as a tool for studying genome-wide variation: A case study of pollinator-mediated divergence in a monkeyflower.' *Molecular Ecology* 26(1): 107–122.

Stebbins, G. L. 1970. 'Adaptive radiation of reproductive characteristics in angiosperms, I: pollination mechanisms.' *Annual Review of Ecology and Systematics* 1(1): 307–26.

Stern, D. L. 2000. 'Perspective: evolutionary developmental biology and the problem of variation.' *Evolution* 54(4): 1079–91.

Stern, D. L. 2011. *Evolution, Development and the Predictable Genome.* Roberts and Company, Greenwood Village, CO.

Stern, D. L., and Orgogozo, V. 2008. 'The loci of evolution: how predictable is genetic evolution?' *Evolution* 62(9): 2155–77.

Stewart, R. N., Asen, S., Massie, D. R., and Norris, K. H. 1979. 'The identification of poinsettia cultivars by HPLC analysis of their anthocyanin content.' *Biochemical Systematics and Ecology* 7(4): 281–87.

Stewart, R. N., Asen, S., Massie, D. R. and Norris, K. H. 1980. 'The identification of poinsettia cultivars by HPLC analysis of their flavonol content.' *Biochemical Systematics and Ecology* 8(2): 119–25.

Stracke, R., Werber, M., and Weisshaar, B. 2001. 'The R2R3-MYB gene family in *Arabidopsis thaliana*.' *Current Opinion in Plant Biology* 4(5): 447–56.

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., and Weisshaar, B. 2007. 'Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling.' *The Plant Journal* 50(4): 660–77.

Stracke, R., Werber, M., and Weisshaar, B. 2001. 'The R2R3-MYB gene family in *Arabidopsis thaliana*.' *Current Opinion in Plant Biology* 4(5): 447–56.

Strathmann, A., Kuhlmann, M., Heinekamp, T., and Dröge-Laser, W. 2001. 'BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development.' *The Plant Journal* 28(4):397-408.

Streisfeld, M. A., and Kohn, J. R. 2007. 'Environment and pollinator-mediated selection on parapatric floral races of *Mimulus aurantiacus*.' *Journal of Evolutionary Biology* 20(1): 122–32.

Streisfeld, M. A., and Rausher, M. D. 2009. 'Altered trans-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*.' *Molecular Biology and Evolution* 26(2): 433–44.

Streisfeld, M. A., and Rausher, M. D. 2011. 'Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution.' *Evolution* 65(3): 629–42.

Streisfeld, M. A., Young, W. N., and Sobel, J. M. 2013. 'Divergent selection drives genetic differentiation in an R2R3-MYB transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*.' *PLoS Genetics* 9(3): e1003385.

Struck, M. 1994. 'Flowers and their insect visitors in the arid winter rainfall region of southern Africa: Observations on Permanent Plots. Insect Visitation Behaviour.' *Journal of Arid Environments* 28(1): 51–74.

Sun, M., Jia, B., Yang, J., Cui, N., Zhu, Y., and Sun, X. 2017. 'Genome-wide identification of the PHD-finger family genes and their responses to environmental stresses in *Oryza sativa* L.' *International Journal of Molecular Sciences* 18(9).

Sun, X., Wang, Y., and Sui, N. 2018. 'Transcriptional regulation of bHLH during plant response to stress.' *Biochemical and Biophysical Research Communications* 503(2): 397–401.

Suzuki, K., Suzuki, T., Nakatsuka, T., Dohra, H., Yamagishi, M., Matsuyama, K., and Matsuura, H. 2016. 'RNA-seq-based evaluation of bicolor tepal pigmentation in Asiatic hybrid lilies (*Lilium* Spp.).' *BMC Genomics* 17(1): 1-19.

Taguchi, G., Shitchi, Y., Shirasawa, S., Yamamoto, H., and Hayashida, N., 2005. 'Molecular cloning, characterization, and downregulation of an acyltransferase that catalyzes the malonylation of flavonoid and naphthol glucosides in tobacco cells.' *The Plant Journal* 42(4): 481–91.

Takeda, K., Harborne, J. B., and Self, R. 1986. 'Identification and distribution of malonated anthocyanins in plants of the Compositae.' *Phytochemistry* 25(6): 1337–42.

Tanaka, Y., Sasaki, N., and Ohmiya, A. 2008. 'Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids.' *The Plant Journal* 54(4): 733–49.

Tang, Y., Horikoshi, M., and Li, W. 2016. 'ggfortify: Unified interface to visualize statistical result of popular R packages.' *The R Journal* (8). https://journal.r-project.org/.

Tateishi, N., Ozaki, Y., and Okubo, H. 2010. 'White marginal picotee formation in the petals of *Camellia japonica* 'Tamanoura.'' *Journal of the Japanese Society for Horticultural Science* 79(2): 207–14.

Taura, H. M., and Laroca, S. 2004. 'Biologia Da Polinização: Interações Entre as Abelhas (Hym., Apoidea) e as Flores de Vassobia Breviflora (Solanaceae).' *Acta Biológica Paranaense* 33: 143–62.

Taylor, S. A., Hofer, J. M., Murfet, I. C., Sollinger, J. D., Singer, S. R., Knox, M. R., and Ellis, T.N. 2002. 'Proliferating inflorescence meristem, a MADS-box gene that regulates floral meristem identity in pea.' *Plant Physiology* 129(3): 1150–59.

Teh, L. S., and Francis, F. J. 1988. 'A research note stability of anthocyanins from *Zebrina pendula* and

Ipomoea tricolor in a model beverage.' *Journal of Food Science* 53(5): 1580–81.

Terahara, N., Callebaut, A., Ohba, R., Nagata, T., Ohnishi-Kameyama, M., and Suzuki, M. 2001. 'Acylated anthocyanidin 3-sophoroside-5-glucosides from *Ajuga reptans* flowers and the corresponding cell cultures.' *Phytochemistry* 58(3): 493–500.

Theißen, G. 2001. 'Development of floral organ identity: stories from the MADS house.' *Current Opinion in Plant Biology* 4(1): 75–85.

Theißen, G., and Saedler, H. 1999. 'The golden decade of molecular floral development (1990-1999): a cheerful obituary.' *Developmental Genetics* 25(3): 181–93.

Theißen, G., and Saedler, H. 2001. 'Floral Quartets.' *Plant biology* 20(11): 2705–7.

Theißen, G., Melzer, R., and Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* 143(18): 3259-3271.

Thomas, M. M., Rudall, P. J., Ellis, A. G., Savolainen, V., and Glover, B. J. 2009. 'Development of a complex floral trait: the pollinator-attracting petal spots of the beetle daisy, *Gorteria diffusa* (Asteraceae).' *American Journal of Botany* 96(12): 2184–96.

Thomas, M. M. 2009. *Evolution of the 'Beetle Daisy' Petal Spot*. University of Cambridge. PhD thesis.

Thomson, F. J., Moles, A. T., Auld, T. D., and Kingsford, R. T. 2011. 'Seed dispersal distance is more strongly correlated with plant height than with seed mass.' *Journal of Ecology* 99(6): 1299–1307.

Tian, J., Chen, M. C., Zhang, J., Li, K. T., Song, T. T., Zhang, X. and Yao, Y. C. 2017. 'Characteristics of d*ihydroflavonol 4-reductas*e gene promoters from different leaf colored malus crabapple cultivars.' *Horticulture Research* 4(1): 1–10.

Trigueros, M., Navarrete-Gómez, M., Sato, S., Christensen, S.K., Pelaz, S., Weigel, D., Yanofsky, M.F., and Ferrándiz, C. 2009. 'The NGATHA genes direct style development in the *Arabidopsis* gynoecium.' *The Plant Cell* 21(5): 1394-1409.

Tripp, E. A., Zhuang, Y., Schreiber, M., Stone, H., and Berardi, A. E. 2018. 'Evolutionary and ecological drivers of plant flavonoids across a large latitudinal gradient.' *Molecular Phylogenetics and Evolution* 128: 147–61.

Turelli, M., Barton, N. H., and Coyne, J. A. 2001. 'Theory of speciation.' *Trends in Ecology and Evolution* 16(7): 330–43.

Turing, A. M. 1953. 'The chemical basis of morphogenesis.' *Bulletin of Mathematical Biology* 52(1): 153–97.

Twyford, A. D., Streisfeld, M. A., Lowry, D. B., and Friedman, J. 2015. 'Genomic studies on the nature of species: adaptation and speciation in *Mimulus*.' *Molecular Ecology* 24(11): 2601–9.

Tyson, P. D., and Partridge, T. C. 2000. 'Evolution of Cenozoic climates.' In *Oxford Monographs on Geology and Geophysics*, 371–87.

Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., Nishino, T., Kusunoki, M., and Nakayama, T. 2007. 'Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis.' *Journal of Biological Chemistry* 282(21): 15812–22.

Ushimaru, A., Watanabe, T., and Nakata, K. 2007. 'Colored floral organs influence pollinator behavior and pollen transfer in *Commelina communis* (Commelinaceae).' *American Journal of Botany*

94(2): 249–58.

Valente, L. M., Manning, J. C., Goldblatt, P., and Vargas, P. 2012. 'Did pollination shifts drive diversification in southern African *Gladiolus* evaluating the model of pollinator-driven speciation.' *American Naturalist* 180(1): 83–98.

Vamosi, Jana C. et al. 2018. 'Macroevolutionary patterns of flowering plant speciation and extinction.' *Annual Review of Plant Biology* 69(1): 685–706.

Vekemans, X., and Hardy, O. J. 2004. 'New insights from fine-scale spatial genetic structure analyses in plant populations.' *Molecular Ecology* 13(4): 921–35.

Verboom, G. A., Linder, H. P., Forest, F., Hoffmann, V., Bergh, N. G., Cowling, R. M., Allsopp, N., and Colville, J. F. 2014. 'Cenozoic assembly of the Greater Cape flora.' In *Fynbos*, Oxford: Oxford University Press, 93–118.

Verçoza, F. C., Dias, A. R., and Missagia, C. C. 2012. 'Ecologia da polinização e potenciais dispersores da ' Marianeira ' - *Acnistus arborescens* (L .) Schltdl . (Solanaceae) Em Área de Floresta Atlântica Do Rio de Janeiro.' *Natureza online* 10(2): 59–64.

Vereecken, N. J., Wilson, C. A., Hötling, S., Schulz, S., Banketov, S. A., and Mardulyn, P. 2012. 'Pre-adaptations and the evolution of pollination by sexual deception: Cope's rule of specialization revisited.' *Proceedings of the Royal Society B: Biological Sciences* 279(1748): 4786–94.

Vittoz, P., and Engler, R. 2007. 'Seed dispersal distances: A typology based on dispersal modes and plant traits.' *Botanica Helvetica* 117(2): 109–24.

Vlasáková, B., Kalinová, B., Gustafsson, M. H. G., and Teichert, H. 2008. 'Cockroaches as pollinators of *Clusia aff. sellowiana* (Clusiaceae) on inselbergs in French Guiana.' *Annals of Botany* 102(3): 295–304.

Vlašánková, A., Padyšáková, E., Bartoš, M., Mengual, X., Janečková, P., and Janeček, Š. 2017. 'The nectar spur is not only a simple specialization for long-proboscid pollinators.' *New Phytologist* 215(4): 1574–81.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., and Giovannoni, J. 2002. 'A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (Rin) locus.' *Science* 296(5566): 343–46.

de Waal, C., Rodger, J. G., Anderson, B., and Ellis, A. G. 2014. 'Selfing ability and dispersal are positively related, but not affected by range position: A multispecies study on southern African asteraceae.' *Journal of Evolutionary Biology* 27(5): 950–59.

de Waal, C., Anderson, B., and Ellis, A. G. 2015. 'Relative density and dispersion pattern of two southern African Asteraceae affect fecundity through heterospecific interference and mate availability, not pollinator visitation rate.' *Journal of Ecology* 103(2): 513–25.

Wagner, C. E., Harmon, L. J., and Seehausen, O. 2012. 'Ecological opportunity and sexual selection together predict adaptive radiation.' *Nature* 487(7407): 366–69.

Wagner, D. 2009. 'Flower morphogenesis: timing is key.' *Developmental Cell* 16(5): 621–22.

Walcher-Chevillet, C. L., and Kramer, E. M. 2016. 'Breaking the mold: understanding the evolution and development of lateral organs in diverse plant models.' *Current Opinion in Genetics and Development* 39: 79–84.

Walker, A. R., Lee, E., Bogs, J., McDavid, D. A., Thomas, M. R., and Robinson, S. P. 2007. 'White grapes arose through the mutation of two similar and adjacent regulatory genes.' *The Plant Journal*

49(5): 772–85.

Walker, R. H. 2012. ‘Determining the regulators of petal spot development in *Gorteria diffusa*.’ University of Cambridge. PhD thesis.

Wallace, J. R. 2012. ‘Imap: Interactive mapping.’ https://cran.r-project.org/package=Imap.

Wang, J. R. 2012. 'On the measurements of genetic differentiation among populations.' *Genetics Research* 94(5) 275-289.

Wang, N., Qu, C., Jiang, S., Chen, Z., Xu, H., Fang, H., Su, M., Zhang, J., Wang, Y., Liu, W., and Zhang, Z. 2018. ‘The proanthocyanidin-specific transcription factor MdMYBPA1 initiates anthocyanin synthesis under low-temperature conditions in red-fleshed apples.’ *The Plant Journal* 96(1): 39–55.

Wang, Y., Zhang, Y., Zhou, R., Dossa, K., Yu, J., Li, D., Liu, A., Mmadi, M. A., Zhang, X. and You, J. 2018. ‘Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame.’ *PLoS ONE* 13(7): 1–21.

War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S. and Sharma, H. C. 2012. ‘Mechanisms of plant defense against insect herbivores.’ *Plant Signaling and Behavior* 7(10): 1306-1320.

Weber, M. G., and Agrawal, A. A. 2014. ‘Defense mutualisms enhance plant diversification.’ *Proceedings of the National Academy of Sciences of the United States of America* 111(46): 16442–47.

Weberling, F. 1992. *Weberling, Focko. Morphology of Flowers and Inflorescences*. Cambridge: Cambridge University Press.

Wei, W., Huang, J., Hao, Y. J., Zou, H. F., Wang, H. W., Zhao, J. Y., Liu, X. Y., Zhang, W. K., Ma, B., Zhang, J. S., and Chen, S. Y. 2009. ‘Soybean GmPHD-type transcription regulators improve stress tolerance in transgenic *Arabidopsis* plants.’ *PLoS One* 4(9): p.e7209.

Wei, M., Liu, A., Zhang, Y., Zhou, Y., Li, D., Dossa, K., Zhou, R., Zhang, X., and You, J. 2019. ‘Genome-wide characterization and expression analysis of the HD-Zip gene family in response to drought and salinity stresses in sesame.’ *BMC Genomics* 20(1): 1–13.

Weir, B. S., and Cockerham, C. C. 1984. ‘Estimating F-statistics for the analysis of population structure.’ *Evolution* 38(6): 1358.

Weng, H., Yoo, C. Y., Gosney, M. J., Hasegawa, P. M., and Mickelbart, M. V. 2012. ‘Poplar GTL1 is a Ca 2+/calmodulin-binding transcription factor that functions in plant water use efficiency and drought tolerance.’ *PLoS One* 7(3): e32925.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J. and Kuhn, M. 2019. ‘Welcome to the tidyverse.’ *Journal of Open Source Software* 4(43): 1686.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. 2005. ‘A note on exact tests of Hardy-Weinberg equilibrium.’ *American Journal of Human Genetics* 76(5): 887–93.

Wilmouth, R. C., Turnbull, J. J., Welford, R. W., Clifton, I. J., Prescott, A. G., and Schofield, C. J. 2002. ‘Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*.’ *Structure*

10(1): 93–103.

Winkel-Shirley, B. 2001. 'Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology.' *Plant Physiology* 126(2): 485–93.

Winter, C. M., Yamaguchi, Wu, M.F., and Wagner, D. 2015. 'Transcriptional programs regulated by both LEAFY and APETALA1 at the time of flower formation.' *Physiologia Plantarum* 155(1): 55–73.

Wittkopp, P. J., and Kalay, G. 2012. 'Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence.' *Nature Reviews Genetics* 13(1): 59–69.

van Wordragen, M. F., de Jong, J., Huitema, H. B., and Dons, H. J., 1991. 'Genetic transformation of chrysanthemum using wild type *Agrobacterium* strains; strain and cultivar specificity.' *Plant Cell Reports* 9(9): 505-508.

Woźniak, N. J., and Sicard, A. 2018. 'Evolvability of flower geometry: convergence in pollinator-driven morphological evolution of flowers.' *Seminars in Cell and Developmental Biology* 79: 3–15.

Wray, G. A. 2007. 'The evolutionary significance of *cis*-regulatory mutations.' *Nature Reviews Genetics* 8(3): 206–16.

Wright, S. 1943. 'Isolation by distance.' *Genetics* 28(2): 114–38.

Wright, S. I., Ness, R. W., Foxe, J. P., and Barretty, S. C. H. 2008. 'Genomic consequences of outcrossing and selfing in plants.' *International Journal of Plant Sciences* 169(1): 105–18.

Wu, C. A., Lowry, D. B., Cooley, A. M., Wright, K. M., Lee, Y. W., and Willis, J. H. 2008. '*Mimulus* is an emerging model system for the integration of ecological and genomic studies.' *Heredity* 100: 220–30.

Wuertz, D., Setz, T., and Chalabi, Y. 2017. 'FOptions: Rmetrics - pricing and evaluating basic options.' *R package version*, 3042.

Xi, J., Qiu, Y., Du, L., and Poovaiah, B. W. 2012. 'Plant-specific trihelix transcription factor AtGT2L interacts with calcium/calmodulin and responds to cold and salt stresses.' *Plant Science* 185: 274-280.

Xia, F., Sun, T., Yang, S., Wang, X., Chao, J., Li, X., Hu, J., Cui, M., Liu, G., Wang, D., and Sun, Y. 2019. 'Insight into the B3 transcription factor superfamily and expression profiling of B3 genes in axillary buds after topping in tobacco (*Nicotiana tabacum* L.).' *Genes* 10(2): 1–28.

Xia, C., Wang, Y. J., Liang, Y., Niu, Q. K., Tan, X. Y., Chu, L. C., Chen, L. Q., Zhang, X. Q., and Ye, D., 2014. 'The ARID-HMG DNA-binding protein AtHMGB15 is required for pollen tube growth in *Arabidopsis thaliana*.' *The Plant Journal* 79(5): 741-756.

Xie, D. Y., Sharma, S. B., Paiva, N. L., Ferreira, D., and Dixon, R. A. 2003. 'Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis.' *Science* 299(5605): 396–99.

Xie, Z. M., Zou, H. F., Lei, G., Wei, W., Zhou, Q. Y., Niu, C. F., Liao, Y., Tian, A. G., Ma, B., Zhang, W. K., and Zhang, J. S. 2009. 'Soybean trihelix transcription factors GmGT-2A and GmGT-2B improve plant tolerance to abiotic stresses in transgenic Arabidopsis.' *PloS one* 4(9): e6898.

Xu, Y., Zong, W., Hou, X., Yao, J., Liu, H., Li, X., Zhao, Y., and Xiong, L., 2015. 'Os ARID 3, an AT-rich interaction domain-containing protein, is required for shoot meristem development in rice.' *The Plant Journal* 83(5): 806-817.

Yamagishi, M. 2016. 'A novel R2R3-MYB transcription factor regulates light-mediated floral and

vegetative anthocyanin pigmentation patterns in *Lilium regale*.' *Molecular Breeding* 36(1): 1–14.

Yamagishi, M. 2018. 'Involvement of a LhMYB18 transcription factor in large anthocyanin spot formation on the flower tepals of the Asiatic hybrid lily *(Lilium* Spp.) cultivar 'Grand Cru.'' *Molecular Breeding* 38(5): 1-16.

Yamagishi, M., Shimoyamada, Y., Nakatsuka, T., and Masuda, K. 2010. 'Two R2R3-MYB genes, homologs of Petunia *AN2*, regulate anthocyanin biosyntheses in flower tepals, tepal spots and leaves of Asiatic hybrid lily.' *Plant and Cell Physiology* 51(3): 463–74.

Yamagishi, M., Toda, S., and Tasaki, K. 2014. 'The novel allele of the *LhMYB12* gene is involved in splatter-type spot formation on the flower tepals of Asiatic hybrid lilies (*Lilium* Spp.).' *New Phytologist* 201(3): 1009–20.

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K. and Yokoyama, S. 2013. 'DNA-binding domains of plant-specific transcription factors: structure, function, and evolution.' *Trends in Plant Science* 18(5): 267–76.

Yang, C. Q., Fang, X., Wu, X. M., Mao, Y. B., Wang, L. J., and Chen, X. Y. 2012. 'Transcriptional regulation of plant secondary metabolism.' *Journal of Integrative Plant Biology* 54(10): 703–12.

Yoder, J. B., Gomez, G., and Carlson, C. J. 2020. 'Zygomorphic flowers have fewer visitors.' *BioRxiv* 743872.

Yong, Y., Zhang, Y., and Lyu, Y. 2019. 'A Myb-related transcription factor from *Lilium lancifolium* l. (LLMYB3) is involved in anthocyanin biosynthesis pathway and enhances multiple abiotic stress tolerance in *Arabidopsis thaliana*.' *International Journal of Molecular Sciences* 20(13): 319.

Yoshida, K., Kitahara, S., Ito, D., and Kondo, T. 2006. 'Ferric ions involved in the flower color development of the Himalayan blue poppy, *Meconopsis grandis*.' *Phytochemistry* 67(10): 992–98.

Yoshida, K., and Negishi, T. 2013. 'The identification of a vacuolar iron transporter involved in the blue coloration of cornflower petals.' *Phytochemistry* 94: 60–67.

Yoshioka, Y., Ohashi, K., Konuma, A., Iwata, H., Ohsawa, R., and Ninomiya, S., 2007. 'Ability of bumblebees to discriminate differences in the shape of artificial flowers of *Primula sieboldii* (Primulaceae).' *Annals of Botany* 99(6): 1175–82.

Youssef, H. M., Eggert, K., Koppolu, R., Alqudah, A. M., Poursarebani, N., Fazeli, A., Sakuma, S., Tagiri, A., Rutten, T., Govind, G., and Lundqvist, U. 2017. 'VRS2 regulates hormone-mediated inflorescence patterning in barley.' *Nature Genetics* 49(1): 157-161.

Yu, X., Chen, M., and Chang-Jun, L. 2008. 'Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7- O- glycosidic (Iso) flavones in *Medicago truncatula*.' *The Plant Journal* 55(3): 382–96.

Yuan, X., Wang, H., Cai, J., Li, D., and Song, F., 2019. 'NAC transcription factors in plant immunity.' *Phytopathology Research* 1(1): 1–13.

Yuan, Y. 2019. 'Monkeyflowers (*Mimulus*): New model for plant developmental genetics and evo-devo.' *New Phytologist* 222(2): 694–700.

Yuan, Y. W., Rebocho, A. B., Sagawa, J. M., Stanley, L. E., and Bradshaw, H. D. 2016. ' Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species .' *Proceedings of the National Academy of Sciences of the United States of America* 113(9): 2448–53.

Yuan, Y. W., Sagawa, J. M., Young, R. C. 2013a. 'Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*.' *Genetics* 194(1): 255–63.

Yuan, Y. W., Sagawa, J. M., Frost, L., Vela, J. P., and Bradshaw Jr, H. D. 2014. 'Transcriptional control of floral anthocyanin pigmentation in monkeyflowers (*Mimulus*).' *New Phytologist* 204(4): 1013–27

Yuan, Y. W., Sagawa, J. M., Di Stilio, V. S., and Bradshaw, H. D. 2013b. 'Bulk segregant analysis of an induced floral mutant identifies a MIXTA-like R2R3 MYB controlling nectar guide formation in *Mimulus lewisii*.' *Genetics* 194(2): 523–28.

Yue, J., Zhu, C., Zhou, Y., Niu, X., Miao, M., Tang, X., Chen, F., Zhao, W., and Liu, Y. 2018. 'Transcriptome analysis of differentially expressed unigenes involved in flavonoid biosynthesis during flower development of *Chrysanthemum morifolium* 'Chuju.'' *Scientific Reports* 8(1): 1–14.

Zhang, J. 2003. 'Evolution by gene duplication: An update.' *Trends in Ecology and Evolution* 18(6): 292–98.

Zhang, B., Wang, L., Zeng, L., Zhang, C., and Ma, H., 2015. '*Arabidopsis* TOE proteins convey a photoperiodic signal to antagonize CONSTANS and regulate flowering time.' *Genes and Development* 29(9): 975-987.

Zhang, T., Zhao, Y., Wang, Y., Liu, Z., and Gao, C. 2018. 'Comprehensive analysis of Myb gene family and their expressions under abiotic stresses and hormone treatments in *Tamarix hispida*.' *Frontiers in Plant Science* 9: 1303.

Zhao, C. L., Yu, Y. Q., Chen, Z. J., Wen, G. S., Wei, F. G., Zheng, Q., Wang, C. D., and Xiao, X. L. 2017. 'Stability-increasing effects of anthocyanin glycosyl acylation.' *Food Chemistry* 214: 119–28.

Zhao, D., and Tao, J. 2015. 'Recent advances on the development and regulation of flower color in ornamental plants.' *Frontiers in Plant Science* 6: 1–13.

Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H., and Li, G. 2017. 'The *Arabidopsis thaliana* nuclear factor Y transcription factors.' *Frontiers in Plant Science* 7: 1–11.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. 2012. 'A high-performance computing toolset for relatedness and principal component analysis of SNP data.' *Bioinformatics* 28(24): 3326–28.

Zheng, Y., Ren, N., Wang, H., Stromberg, A. J., and Perry, S. E. 2009. 'Global identification of targets of the *Arabidopsis* MADS domain protein AGAMOUS-Like15.' *The Plant Cell* 21(9): 2563–77.

Zimmermann, I. M., Heim, M. A, Weisshaar, B., and Uhrig, J. F. 2004. 'Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like bHLH proteins.' *The Plant Journal* 40(1): 22–34.

Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z. and Zhang, Z., 2008. 'A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*.' *Plant Physiology* 148(1): 337-347.

Zhong, S., Zhao, M., Shi, T., Shi, H., An, F., Zhao, Q., and Guo, H. 2009. 'EIN3/EIL1 cooperate with PIF1 to prevent photo-oxidation and to promote greening of Arabidopsis seedlings.' *Proceedings of the National Academy of Sciences of the United States of America* 106(50): 21431-21436.

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

Zufall, R A, and Rausher, M. D. 2003. 'The genetic basis of a flower color polymorphism in the common

morning glory (*Ipomoea purpurea*).' *Journal of Heredity* 94(6): 442–48.

Zvi, M. M. B., Shklarman, E., Masci, T., Kalev, H., Debener, T., Shafir, S., Ovadis, M., and Vainstein, A. 2012. 'PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers.' *New Phytologist* 195(2): 335–45.

Appendices

Appendix 1. Media and Solutions

CTAB
2% (v/v) β-Mercaptoethanol, 2% (w/v) CTAB, 2% (w/v) PVP, 1.4M sodium chloride, 20mM EDTA pH 8, 100mM tris-hydrochloride pH 8

gDNA extraction buffer
250mM Tris-HCL pH 7.5, 250mM NaCl, 25mM EDTA, 1% SDS

TE buffer
10mM Tris-HCl pH 7.6, 1mM EDTA pH8

DNase I (homemade buffer)
10mM tris-hydrochloride, 1mM calcium chloride, and 2.5mM magnesium chloride

TBE buffer
45mM Tris base, 45mM boric acid, 2mM EDTA pH8

Gel loading buffer
100mg Orange G dye, 15ml glycerol, up to 50ml with H2O

LB
10g/l tryptone, 10g/l sodium chloride, 5g/l yeast Extract
For solid media: 6g/l bacto-Agar (Sigma-Aldrich)

E. coli freezing solution
60mM $CaCl_2$, 15% (v/v) glycerol, 10mM PIPES pH7

Eco-Taq Master Mix
0.5µl ecotaq enzyme, 1µl forward primer, 1µl reverse primer, 4µl ecotaq buffer, 13.5µl ddH_2O

Alkaline lysis purification solutions
Solution 1: 50mM glucose, 25mM TrisHCl pH8, 10mM EDTA pH8
Solution 2: 0.2M NaOH, 1% (w/v) SDS
Solution 3: 3M potassium acetate, 5M glacial acetic acid

Gibson assembly isothermal buffers
2% (w/v) PEG-80000, 500mM Tris-HCL pH7.5, 50mM $MgCl_2$, 50mM DTT, 1mM dATP, 1mM dTTP, 1mM dCTP, 1mM dGTP, 5mM NAS

Half-MS
2.2g/l Murashige-Skoog Medium with vitamins (Duchefa), 35g/l sucrose, ddH_2O

MS9 media (tobacco transformation)
4.4g/l Murashige-Skoog Medium with vitamins (Duchefa), 20g/l sucrose, ddH_2O

MS media (tobacco transformation)
containing 4.4g/l Murashige-Skoog Medium with vitamins (Duchefa), 30g/l sucrose, ddH_2O

Appendix 2. Primers

Name	Amplified product description	Sequence	Chapter
GdActin.F	Reference gene for checking cDNA synthesis has worked	CCAAGGGCAGTGTTTCCTAGT	2
GdActin.R	Reference gene for checking cDNA synthesis has worked	TGGTACGACCACTGGCATAG	2
GeneRacer.R	3'RACE primer	GCTGTCAACGATACGCTACGTAACG	2
GeneRacer.Rn	3'RACE nested primer	CGCTACGTAACGGCATGACAG	2
M13.F	Used in colony PCR to amplify insert in the pBLUE vector	GTAAAACGACGGCCAGT	2
M13.R	Used in colony PCR to amplify insert in the pBLUE vector	CAGGAAACAGCTATGAC	2
GdEF-2.qP.F	Reference gene for qRT-PCR in Spring	CAACTGCAGCGGGTCCATTAT	4
GdEF-2.qP.R	Reference gene for qRT-PCR in Spring, Cal, and Stein	CAGCTGTCATAACTTGCCCTGAA	4
GdEF-2C.qP.F	Reference gene for qRT-PCR in Cal and Stein	ACTGCAGCGGGTCCATTATGT	4
GdMYB8a.qP.F	*GdMYB8a* for qRT-PCR in Spring and Stein	TCTACTCAAAACACCAAATGATGATCTT	4
GdMYB8a.qP.R	*GdMYB8a* for qRT-PCR in Spring, Cal, and Stein	AACTTCGGTGCCAGTGTT	4
GdMYB8aC.qP.F	*GdMYB8a* for qRT-PCR in Cal	CCTACTCAAAACACCAAATGATGATCTT	4
GdMYB8b.qP.F	*GdMYB8b* for qRT-PCR in Spring and Cal	AAACACGAAAAGCAAACGGACA	4
GdMYB8b.qP.R	*GdMYB8b* for qRT-PCR in Spring	AAAAGATAGGGTTAGGAACATCAACGT	4
GdMYB8bC.qP.R	*GdMYB8b* for qRT-PCR in Cal	GGTTAGGAACATCAACGGTGTGAA	4
GdMYB8c.qP.F	*GdMYB8c* for qRT-PCR in Spring, Cal, and Stein	TTCTTTGGTGGAGAGGCAGG	4
GdMYB8c.qP.R	*GdMYB8c* for qRT-PCR in Spring, Cal, and Stein	GATAGGGGTAGGAACATCAAAGTTGTT	4
GdMYB8d.qP.F	*GdMYB8d* for qRT-PCR in Spring	ACAAAATGTCCCCAACTTTAATCTCGTC	4
GdMYB8d.qP.R	*GdMYB8d* for qRT-PCR in Spring	GACCACCCAATTTCACAGTCAAATTCATC	4
GdMYB8dC.qP.F	*GdMYB8d* for qRT-PCR in Cal and Stein	GTGGTCATTTGGTGGTTCTTCGA	4
GdMYB8dC.qP.R	*GdMYB8d* for qRT-PCR in Cal and Stein	TTCCGAATGTAGCAAGTCCCACATG	4
GdANS.qP.F	*GdANS* for qRT-PCR in Spring and Cal	CGTTCCCGGAGGAGAAAC	4
GdANS.qP.R	*GdANS* for qRT-PCR in Spring and Cal	GTGGCGAGTGCTCGTAG	4
GdDFR.qP.F	*GdDFR* for qRT-PCR in Spring and Cal	GCGAAAACAGTCAAGAGGCTAGTT	4
GdDFR.qP.R	*GdDFR* for qRT-PCR in Spring and Cal	AATGTCCCTCATCGTAAACAGGAAGT	4
GdMAT1.qP.F	*GdMAT1* for qRT-PCR in Spring, Cal, and Stein	CCATCTTTTCTTCTACGAATTCCCCTACTC	4
GdMAT1.qP.R	*GdMAT1* for qRT-PCR in Spring, Cal, and Stein	TGACACCTGAATTATTAGGGTTGGA	4

Name	Amplified product description	Sequence	Chapter
GdMYB8a.F	Used to amplify full length *GdMYB8a*	GAAATAGAAATGAGCATGTACTTC	4
GdMYB8a.R	Used to amplify full length *GdMYB8a*	CAATCTCAATCTTTATGCAAC	4
GdMYB8b.F	Used to amplify full length *GdMYB8b*	GAATTTCATCTTTGTCTTCTACA	4
GdMYB8b.R	Used to amplify full length *GdMYB8b*	CAACTTCCATTCATCTTGGA	4
GdMYB8c.F	Used to amplify full length *GdMYB8c*	CAATAAAAACGAGCATGTACATC	4
GdMYB8c.R	Used to amplify full length *GdMYB8c*	AATATGAGTAAAAGATAGGGGTAGGA	4
GdMYB8d.F	Used to amplify full length *GdMYB8d*	ACATGATAGGCTCCTCCCATCTG	4
GdMYB8d.R	Used to amplify full length *GdMYB8d*	CTAGAGACAATTATTTCACAGGAATCTGAG	4
NtEF1.qP.F	*NtEF-1* reference gene for qRT-PCR in *N. tabacum*	TGAGATGCACCACGAAGCTC	5
NtEF1.qP.R	*NtEF-1* reference gene for qRT-PCR in *N. tabacum*	CGTTAAACCCAACATTGTCACCAG	5
NtUBC.qP.F	*NtUBC* reference gene for qRT-PCR in *N. tabacum*	GCAGCACGCATGTTCAGTGA	5
NtUBC.qP.R	*NtUBC* reference gene for qRT-PCR in *N. tabacum*	CAGTCTGCTGTCCAGCTCTG	5
GdMYB8a.tob.qP.F	*GdMYB8a* amplifying transgene for qRT-PCR in *N. tabacum*	AGGACAAGACACCACAGTCACA	5
GdMYB8a.tob.qP.F	*GdMYB8a* amplifying transgene for qRT-PCR in *N. tabacum*	CCTGAACCATTATGAACCCATTTAGGT	5
GdMYB8bc.tob.qP.F	*GdMYB8b* and *GdMYB8c* amplifying transgene for qRT-PCR in *N. tabacum*	CACAAGACACCACGGTCACA	5
GdMYB8bc.tob.qP.F	*GdMYB8b* and *GdMYB8c* amplifying transgene for qRT-PCR in *N. tabacum*	TGACCCATCATGAACCCATTTAGGT	5
NtANS.qP.F	*NtANS* gene for qRT-PCR in *N. tabacum*	TCCTCCACAATATGGTGCCTG	5
NtANS.qP.R	*NtANS* gene for qRT-PCR in *N. tabacum*	GGGTGTCCCCAATATGCATGA	5
NtDFR.qP.F	*NtDFR* gene for qRT-PCR in *N. tabacum*	ACTGAGTTTAAAGGCATCGATAAGGACT	5
NtDFR.qP.R	*NtDFR* gene for qRT-PCR in *N. tabacum*	TGAATTGAAACCCCATATCCGTCAG	5
NtMAT.qP.F	*NtMAT* gene for qRT-PCR in *N. tabacum*	CTCCTGATAAGGTTCGAGGTACAT	5
NtMAT.qP.R	*NtMAT* gene for qRT-PCR in *N. tabacum*	ATTCCATTCCATTCTCGTCGATCTCTTC	5
GdANS.F	Used to amplify full length *GdANS*	CACAACAAAACCACAAACAC	5
GdANS.R	Used to amplify full length *GdANS*	CAAAGAGCAACACTAATGTGATG	5
GdMAT1.F	Used to amplify full length *GdMAT1*	CACCATCCTCTCTCAACCAATTCA	5
GdMAT1.R	Used to amplify full length *GdMAT1*	CTCGAAACAAATCAAAACCAATCA	5
GdDFR1.F	In 5' UTR of *GdDFR* to amplify into gene	ACACTCACCACTCACCAGT	5
GdDFR2.F	At start codon of *GdDFR* to amplify into gene	AAATGAAAGAGGATTCTCCTACCAC	5

Name	Amplified product description	Sequence	Chapter
GdDFR1.R	In 3'UTR of one *GdDFR* 'variant' to amplify into gene	CTTGATTTTATTGACTTGAACCA	5
GdDFR2.R	In 3'UTR of several *GdDFR* 'variant' to amplify into gene	TACAAACCCCTGCCACATC	5
GdDFR3.R	In 3'UTR of one *GdDFR* 'variant' (C) to amplify into gene	CACCGTTTGTAATCTTTTCATTTACAG	5
GdDFR4.R	In 3'UTR of one *GdDFR* 'variant' (D) to amplify into gene	GAGCACCATTTGTAATCTTATTATGTAGA	5
GdDFR5.R	In 3'UTR of one *GdDFR* 'variant' (A) to amplify into gene	TGACCATCAACGTTTTTGACAGAA	5
GdDFR6.R	In 3'UTR of all *GdDFR* 'variant' to amplify into gene	GCACCATTTGTAATCTTGTCATTTAGAG	5
Nt.geno.8a.F	Used to genotype *GdMYB8a* transgenic *N. tabacum*	AGGACAAGACACCACAGTCACA	5
Nt.geno.8bc.F	Used to genotype *GdMYB8b* transgenic *N. tabacum*	CACAAGACACCACGGTCACA	5
Nt.geno.R	Used to genotype transgenic *N. tabacum,* in transcribed portion of 35S terminator	TTATCGGGAAACTACTCACACA	5
Y1H.DFR-1.F	*GdDFR* promoter fragment 1 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCCCATCAACCATCATGCATG	6
Y1H.DFR-1.R	*GdDFR* promoter fragment 1 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCTTTGTTTTTTGTGTGTTTTGT	6
Y1H.DFR-2.F	*GdDFR* promoter fragment 2 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCCATCCGAACATGCCGAA	6
Y1H.DFR-2.R	*GdDFR* promoter fragment 2 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCGATGGTTGCATGCATGAT	6
Y1H.DFR-3.F	*GdDFR* promoter fragment 3 and restriction sites for cloning (also used in yeast colony PCR)	GGCACGCGTCGGCATGTTCGGATGTTC	6
Y1H.DFR-3.R	*GdDFR* promoter fragment 3 and restriction sites for cloning (also used in yeast colony PCR)	GGCACGCGTTTCGGCATGTTCGGATG	6
Y1H.DFR-4.F	*GdDFR* promoter fragment 4 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCCTTAATAGCTCATTAACTTATGTG	6
Y1H.DFR-4.R	*GdDFR* promoter fragment 4 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCCATCTAACCAAACAGCTTATACT	6
Y1H.MAT1-1.F	*GdMAT1* promoter fragment 1 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCCGTACAAATATAAATCATAAACATATCAT	6
Y1H.MAT1-1.R	*GdMAT1* promoter fragment 1 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCGAATTGGTTGAGAGAGGAT	6
Y1H.MAT1-2.F	*GdMAT1* promoter fragment 2 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCAGTTCAATATTTTTATATCTCGTT	6
Y1H.MAT1-2.R	*GdMAT1* promoter fragment 2 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCTGTTTATGATTTATATTTGTACGA	6
Y1H.MAT1-3.F	*GdMAT1* promoter fragment 3 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCTGACAATGTTATAAATCTCTTATATG	6
Y1H.MAT1-3.R	*GdMAT1* promoter fragment 3 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCGAGATATAAAAATATTGAACTCTT	6
Y1H.MAT1-4.F	*GdMAT1* promoter fragment 4 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCGGTTAACATGACTTGACTCA	6
Y1H.MAT1-4.R	*GdMAT1* promoter fragment 4 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCGAGATTTATAACATTGTCACATT	6
Y1H.MAT1-5.F	*GdMAT1* promoter fragment 5 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCTTGAAAGGCAATACGTTGTT	6
Y1H.MAT1-5.R	*GdMAT1* promoter fragment 5 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCGTCAAGTCATGTTAACCAATTCA	6
Y1H.MAT1-6.F	*GdMAT1* promoter fragment 6 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCTATATTACCGTGCGTCCAAA	6

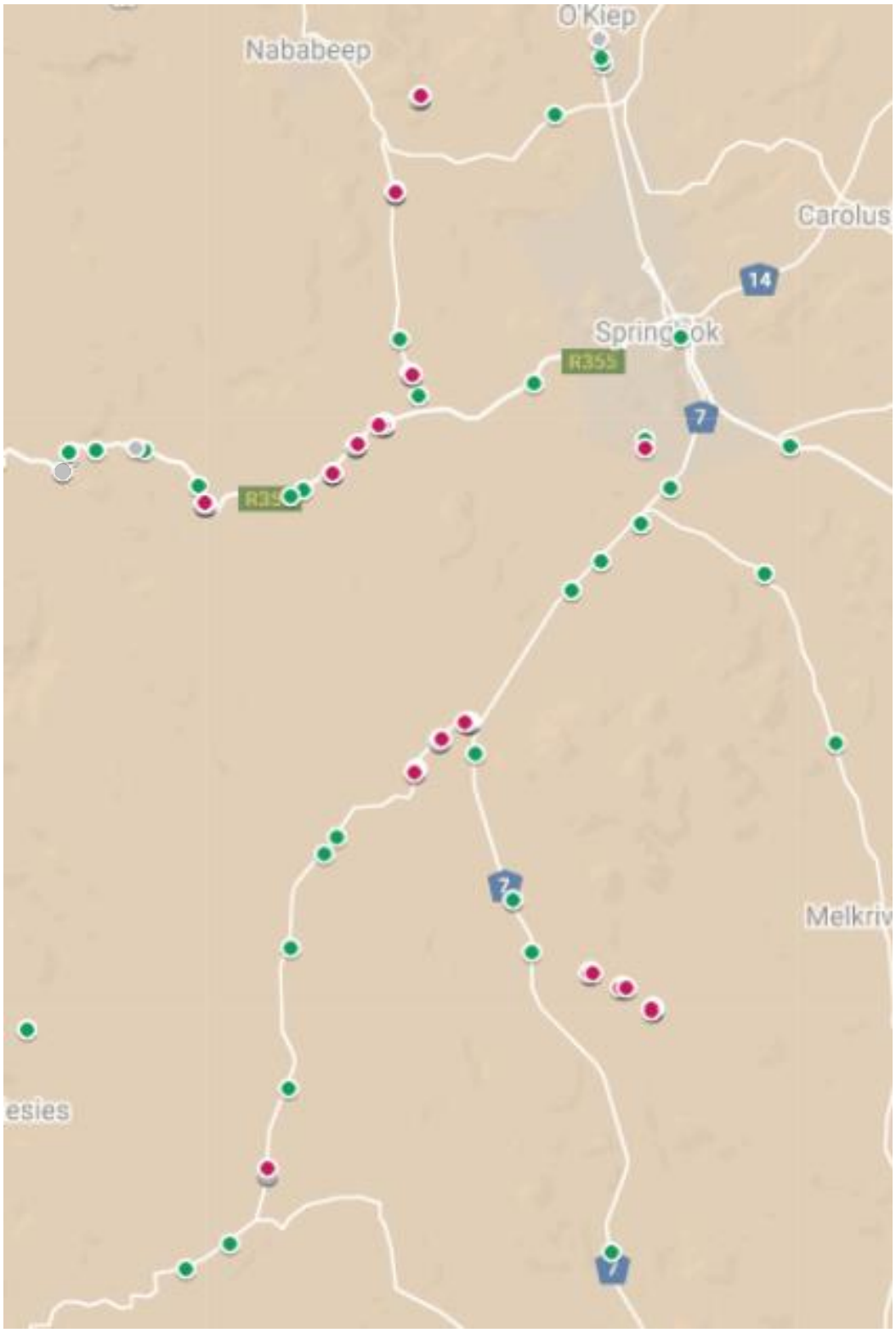
Name	Amplified product description	Sequence	Chapter
Y1H.MAT1-6.R	*GdMAT1* fragment 6 and restriction sites for cloning (used in yeast colony PCR)	GGCGAGCTCCAAAAAATATACTTCTTAAATGTCTAT	6
Y1H.ANS-1.F	*GdANS* fragment 1 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCCATCATCCTCCCATAAAACCA	6
Y1H.ANS-1.R	*GdANS* fragment 1 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCGGATGTATTCTTTTGGGATTTGAT	6
Y1H.ANS-2.F	*GdANS* fragment 2 and restriction sites for cloning (also used in yeast colony PCR)	CCGGAATTCAATGGATATAAAAAGTACCAATATAAGCACA	6
Y1H.ANS-2.R	*GdANS* fragment 2 and restriction sites for cloning (also used in yeast colony PCR)	GGCGAGCTCTGGTTTTATGGGAGGATGATG	6
PR.GdDFR.F	*GdDFR* promoter region	AGAAACCATGTTACTTGTTACGACA	6
PR.GdMAT1.F	*GdMAT1* promoter region	GTAACCGCTTTCTACCTTCTATCTCTCTT	6
PR.GdMYB8a.F	*GdMYB8a* promoter region	CCTATAACTAATTAGAAACACTACCGTAC	6
PR.GdMYB8b.F	*GdMYB8b* promoter region	TCTATTTATTATCTGGAGCTCAACTTT	6
PR.GdMYB8c.F	*GdMYB8c* promoter region	AAGAACCCCAAATCATTTTCCTTT	6
TG.GdMYB8a.F	Used to amplify *GdMYB8a* transgene in *G. diffusa* transformation	AATGTACAACACCACTTGC	6
TG.GdMYB8a.R	Used to amplify *GdMYB8a* transgene in *G. diffusa* transformation	TCAAAGTTGTCCTGAATATAAC	6

Appendix 3. Chapter 3 Supplementary Information

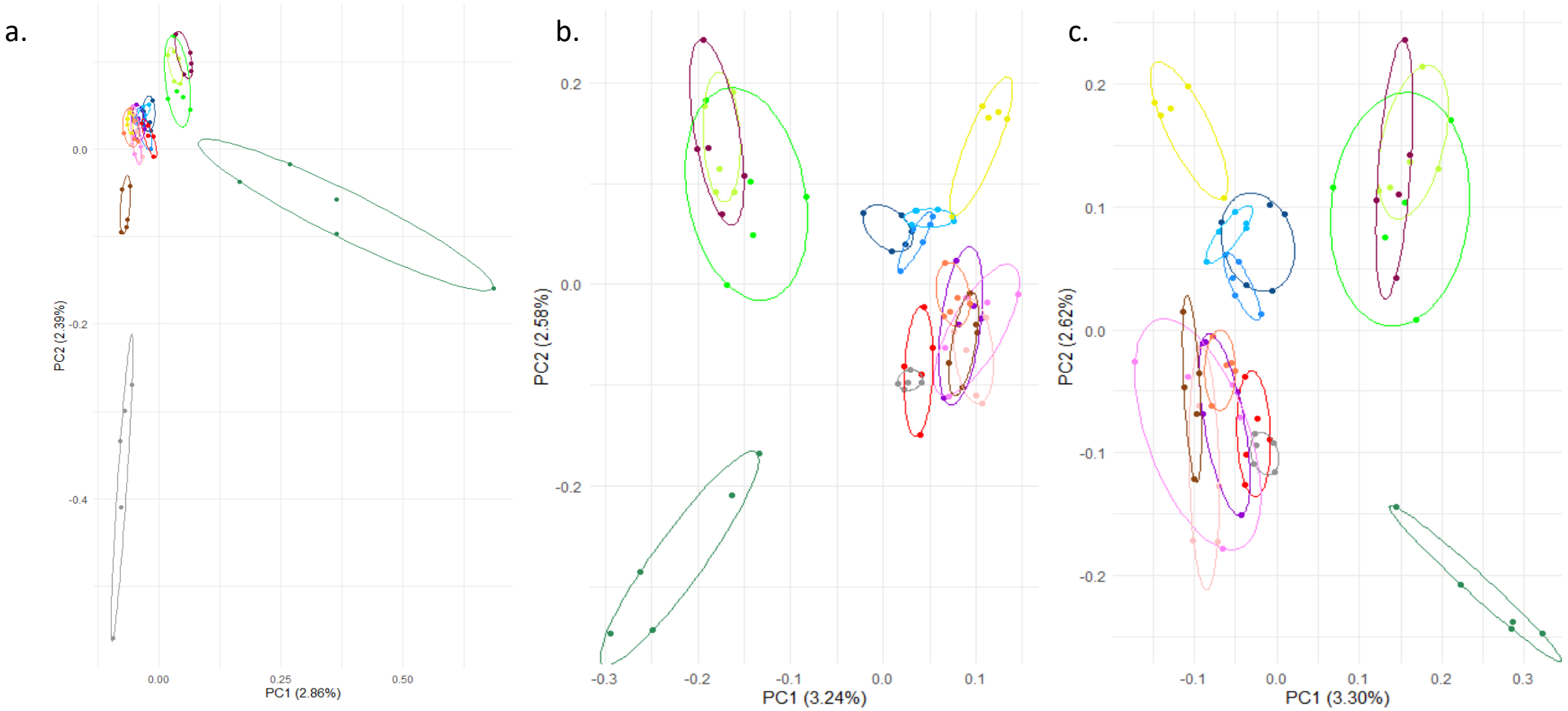
The filtering criteria outline in Section 3.2.2 resulted in a final dataset of 75 samples genotyped with 6,231 SNPs. The retained SNPs for different MAF values were as follows: MAF0 12772 sites kept, mean depth=33.3, stdev=10.3; MAF 0.05 2798 sites kept, mean depth=33.3, stdev=11.2; MAF 0.1 1920 sites kept, mean depth=33.5, stdev=11.3.

Population	Proportion of individuals containing marks	Sample size
H01	0.86	7
H02	0.86	7
H03	0.88	8
H04	1.00	5
M01	0.71	7
M02	0.71	7
M03	0.71	7
R01	1.00	7
R02	1.00	7
SP01	0.58	12
SP02	0.58	12
SP03	0.58	12
NA10	0.58	12
NA11	0.75	12

Supplementary Table 1. The proportion of individuals from each population that have 'marks' - anthocyanin at the base of the plain ray floret.

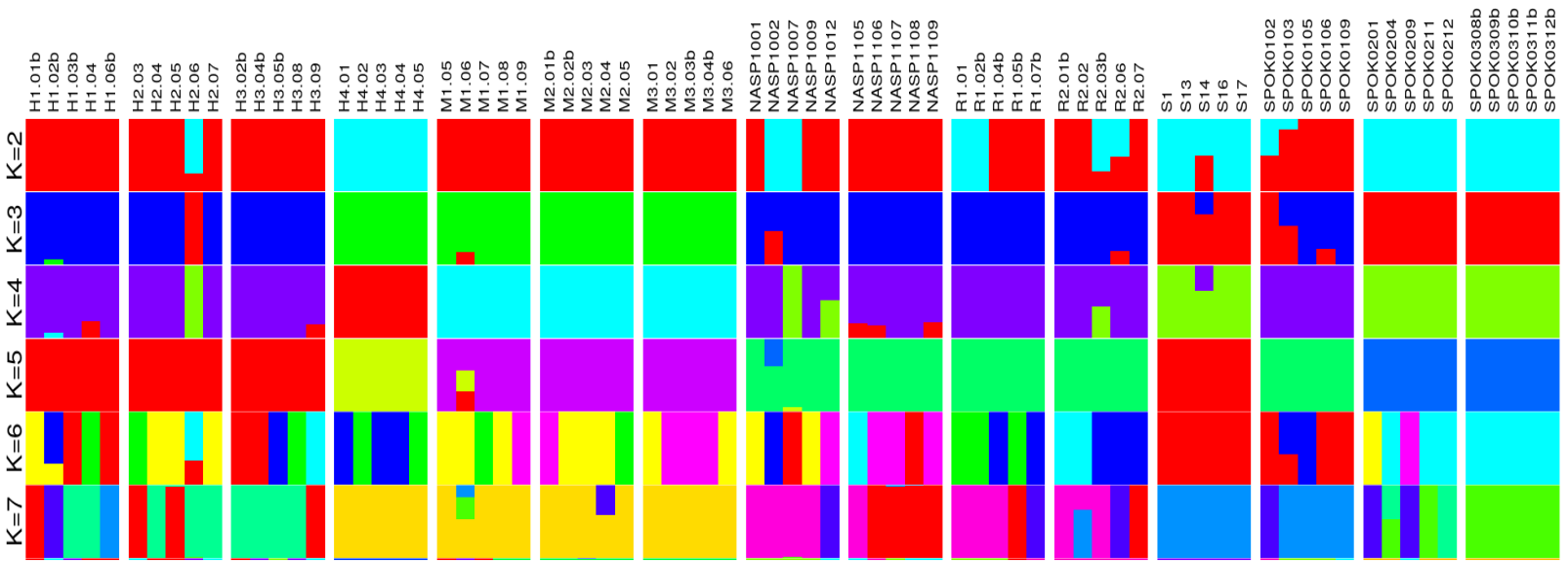


Supplementary Figure 1. Map of the sites where the Spring morphotype of *G. diffusa* has been identified, represented by dots. Dot colour indicates the following: (pink) sites sampled in this analysis, (green) Spring sites not sampled, (grey) contact sites with another morphotype containing hybrids.

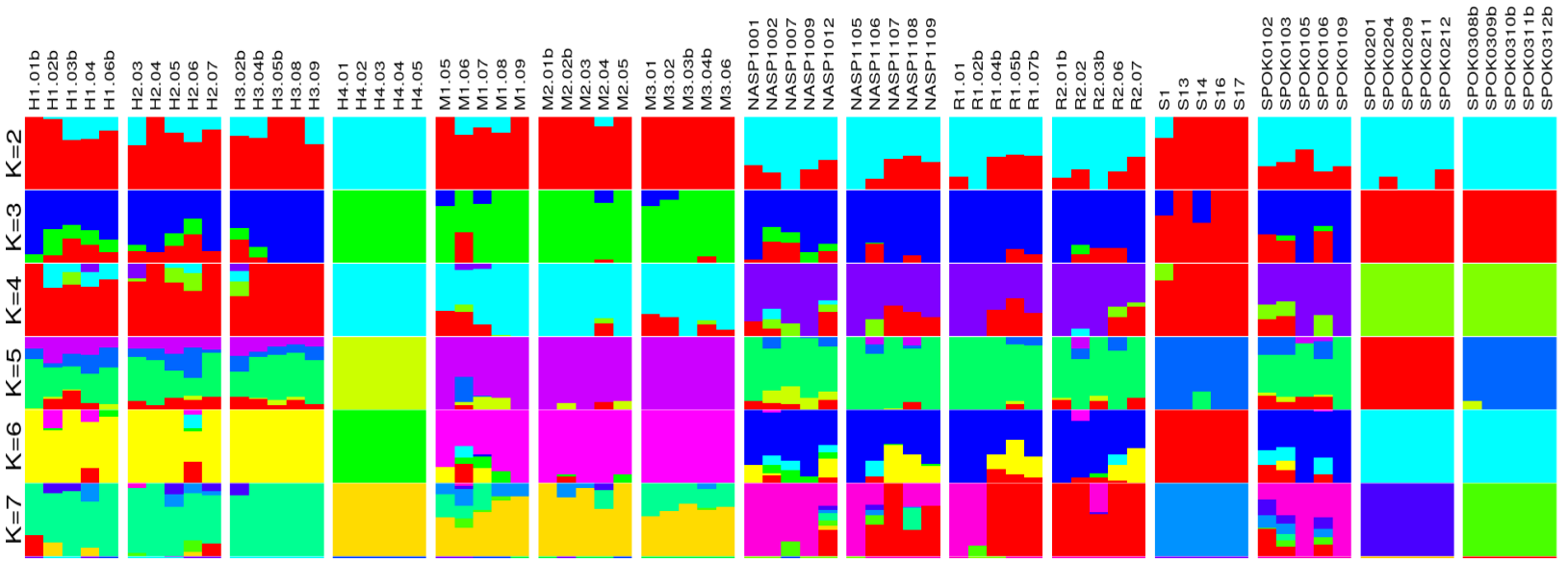


Supplementary Figure 2. The first 2 principal components from a PCA analysis on the genetic structure within the Spring morphotype. Each graph is from a dataset containing a different minimum allele frequency (MAF): a) MAF=0, b) MAF=0.05, c) MAF=0.10

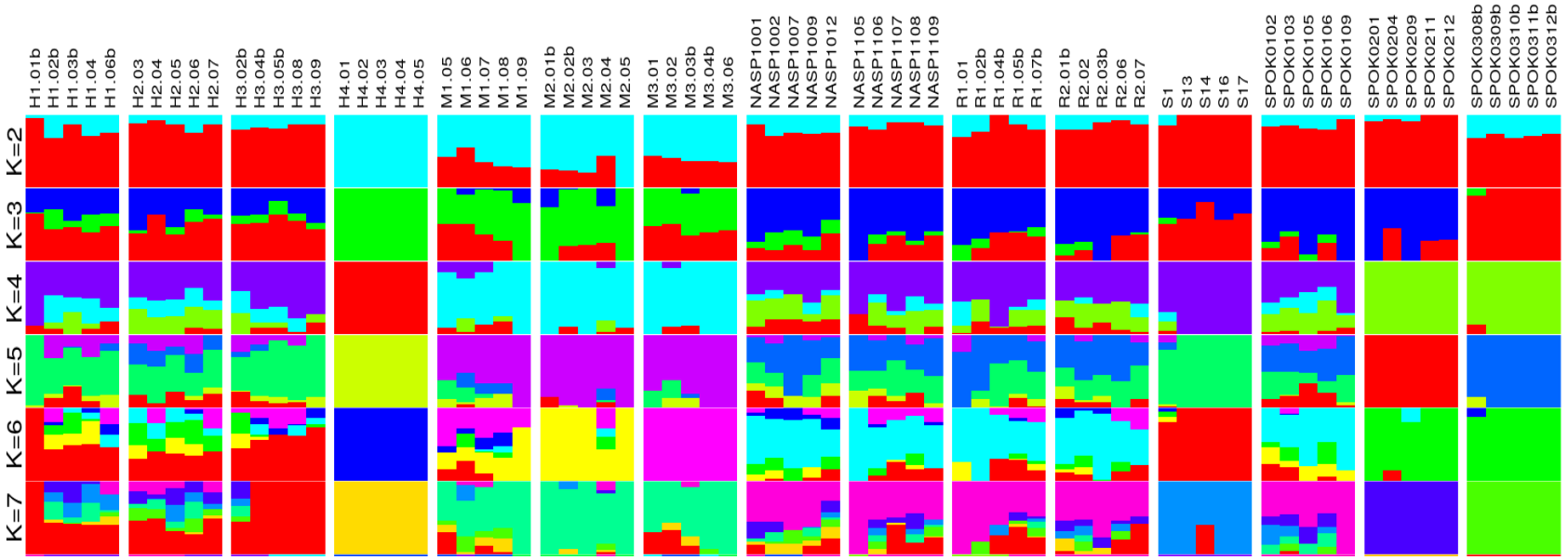
a.


H1.01b H1.02b H1.03b H1.04 H1.06b H2.03 H2.04 H2.05 H2.06 H2.07 H3.02b H3.04b H3.05b H3.08 H3.09 H4.01 H4.02 H4.03 H4.04 H4.05 M1.05 M1.06 M1.07 M1.08 M1.09 M2.01b M2.02b M2.03 M2.04 M2.05 M3.01 M3.02 M3.03b M3.04b M3.06 NASP1001 NASP1002 NASP1007 NASP1009 NASP1012 NASP1105 NASP1106 NASP1107 NASP1108 NASP1109 R1.01 R1.02b R1.04b R1.05b R1.07b R2.01b R2.02 R2.03b R2.06 R2.07 S1 S13 S14 S16 S17 SPOK0102 SPOK0103 SPOK0105 SPOK0106 SPOK0109 SPOK0201 SPOK0204 SPOK0209 SPOK0211 SPOK0212 SPOK0308b SPOK0309b SPOK0310b SPOK0311b SPOK0312b
K=2
K=3
K=4
K=5
K=6
K=7

b.


H1.01b H1.02b H1.03b H1.04 H1.06b H2.03 H2.04 H2.05 H2.06 H2.07 H3.02b H3.04b H3.05b H3.08 H3.09 H4.01 H4.02 H4.03 H4.04 H4.05 M1.05 M1.06 M1.07 M1.08 M1.09 M2.01b M2.02b M2.03 M2.04 M2.05 M3.01 M3.02 M3.03b M3.04b M3.06 NASP1001 NASP1002 NASP1007 NASP1009 NASP1012 NASP1105 NASP1106 NASP1107 NASP1108 NASP1109 R1.01 R1.02b R1.04b R1.05b R1.07b R2.01b R2.02 R2.03b R2.06 R2.07 S1 S13 S14 S16 S17 SPOK0102 SPOK0103 SPOK0105 SPOK0106 SPOK0109 SPOK0201 SPOK0204 SPOK0209 SPOK0211 SPOK0212 SPOK0308b SPOK0309b SPOK0310b SPOK0311b SPOK0312b
K=2
K=3
K=4
K=5
K=6
K=7

C.



Supplementary Figure 3. Admixture plots illustrating the clustering of individuals, based on genetic variants. The x axis represents individuals segregated into populations and the y axis indicates the K value which is the number of clusters specified in the analysis. a) - c) are the same dataset but filtered with different MAF values (a) 0, (b) 0.05, (c) 0.1.

a.

MAF0	H01	H02	H03	H04	M01	M02	M03	NA10	NA11	R01	R02	S	SP01	SP02	SP03
H01		0.0055	0.0065	0.0831	0.029	0.0404	0.0476	0.0328	0.0256	0.0169	0.0258	0.0525	0.0223	0.0482	0.0674
		0.037	0.0285	0.1075	0.0483	0.0621	0.0719	0.0519	0.0476	0.0414	0.0485	0.0778	0.0449	0.0711	0.095
H02	0.019329		0.0025	0.0902	0.0285	0.0372	0.0441	0.025	0.0154	0.0124	0.0243	0.053	0.0213	0.0321	0.0525
			0.0285	0.1208	0.0523	0.0634	0.0723	0.0509	0.0427	0.04	0.0485	0.0826	0.0458	0.0648	0.0792
H03	0.018463	0.015225		0.0839	0.0361	0.0437	0.046	0.0291	0.025	0.0201	0.0278	0.0497	0.024	0.043	0.0621
				0.1126	0.0528	0.0639	0.0708	0.0518	0.0496	0.0421	0.0455	0.071	0.0477	0.0687	0.0856
H04	0.095907	0.103447	0.09812		0.0636	0.0831	0.0896	0.078	0.0919	0.0833	0.0829	0.1338	0.0906	0.1084	0.1057
					0.0867	0.1081	0.1145	0.1033	0.1165	0.108	0.1065	0.1607	0.1153	0.1376	0.1341
M01	0.039344	0.039289	0.043716	0.07446		0.0189	0.0196	0.0361	0.0368	0.0307	0.0342	0.06	0.0224	0.0584	0.0603
						0.0388	0.0372	0.0562	0.0554	0.052	0.0512	0.0786	0.043	0.0825	0.0808
M02	0.051706	0.04924	0.053245	0.09412	0.028798		0.0376	0.0469	0.0454	0.0377	0.0424	0.0709	0.0351	0.0667	0.0769
							0.058	0.0722	0.0645	0.0569	0.0605	0.0943	0.0518	0.0919	0.102
M03	0.058091	0.05727	0.055187	0.10108	0.028857	0.04702		0.0583	0.0501	0.0425	0.0523	0.0823	0.0505	0.0695	0.0757
								0.076	0.0692	0.0634	0.0717	0.1026	0.0691	0.0975	0.0966
NA10	0.041479	0.039023	0.041586	0.088825	0.045875	0.058612	0.066049		0.0236	0.0153	0.0209	0.0602	0.0169	0.0476	0.056
									0.0403	0.0322	0.0368	0.0851	0.0324	0.0685	0.0809
NA11	0.035693	0.029351	0.036472	0.102972	0.044153	0.054829	0.060628	0.029114		0.0007	0.0046	0.0541	0.0099	0.0422	0.0585
										0.0172	0.0206	0.0751	0.0288	0.0651	0.0836
R01	0.029406	0.026568	0.030649	0.097166	0.04082	0.047701	0.054005	0.025108	0.009383		-0.0034	0.039	0.0018	0.0405	0.0537
											0.0129	0.061	0.0212	0.0615	0.0774
R02	0.035572	0.035357	0.035962	0.096804	0.042915	0.051171	0.061565	0.028314	0.011983	0.00438		0.0499	0.0076	0.0333	0.0553
												0.0678	0.0219	0.0544	0.0781
S	0.064774	0.069213	0.059998	0.148732	0.068862	0.082428	0.092634	0.072476	0.064938	0.048559	0.060171		0.0436	0.0708	0.0933
													0.0654	0.096	0.1188
SP01	0.033368	0.0319	0.034857	0.104471	0.03225	0.043286	0.059603	0.024058	0.019805	0.01122	0.015257	0.053376		0.0323	0.0624
														0.0532	0.0853
SP02	0.059389	0.049686	0.057413	0.123966	0.069157	0.078154	0.084004	0.057407	0.054147	0.050874	0.045183	0.083305	0.043052		0.0612
															0.0832
SP03	0.080601	0.064216	0.072487	0.118254	0.070581	0.089198	0.086393	0.068462	0.070565	0.065813	0.066285	0.105215	0.072274	0.074269	

b.

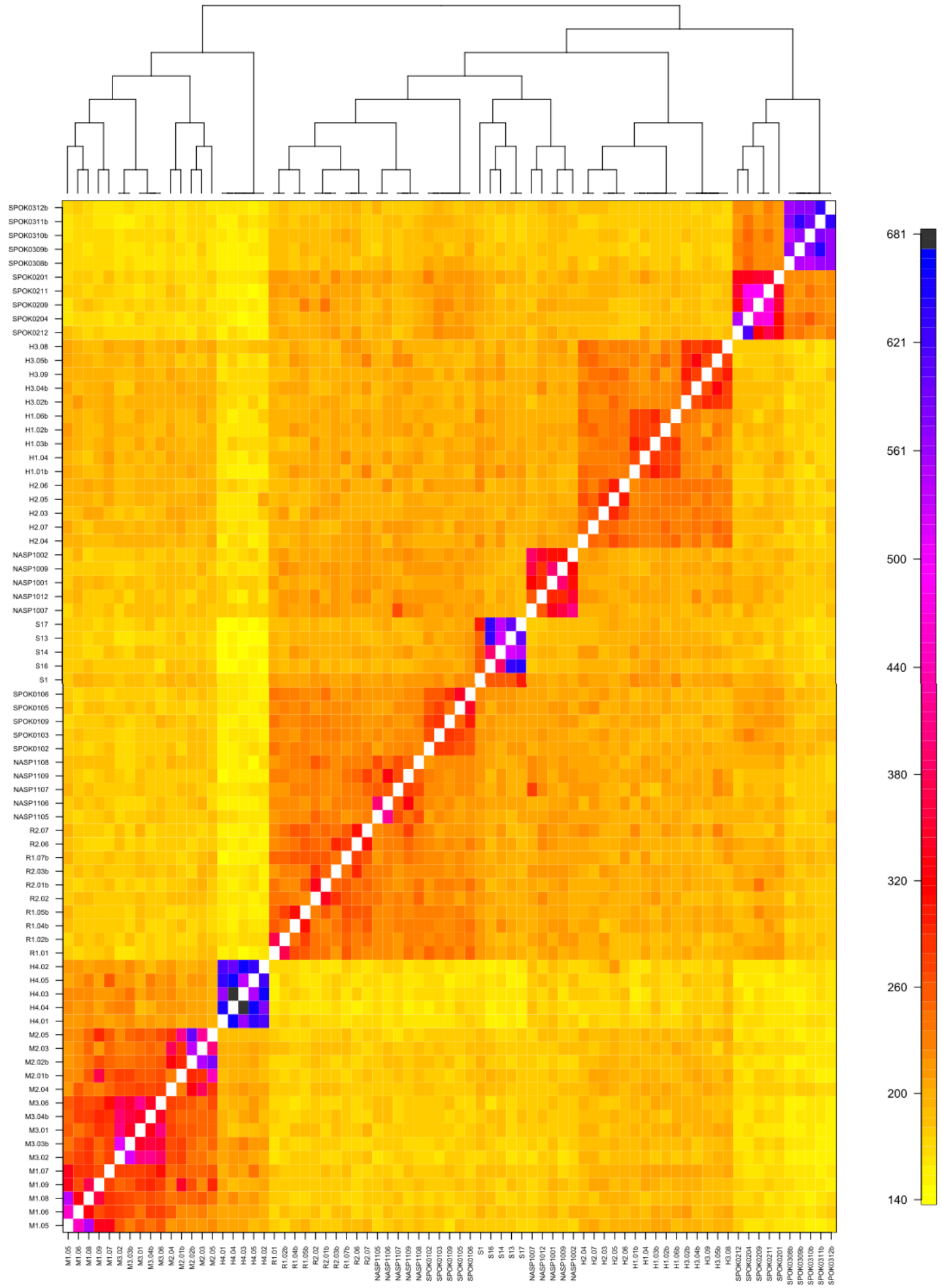
MAF5	H01	H02	H03	H04	M01	M02	M03	NA10	NA11	R01	R02	S	SP01	SP02	SP03
H01		0.01	0.0064	0.081	0.0253	0.0387	0.0447	0.0286	0.0205	0.0148	0.0223	0.0554	0.0212	0.0453	0.0678
		0.042	0.0359	0.1167	0.0491	0.0668	0.0712	0.0572	0.0485	0.043	0.0482	0.0818	0.0462	0.0707	0.0962
H02	0.025797		-0.0019	0.0948	0.0276	0.0363	0.0503	0.0258	0.0155	0.0109	0.0231	0.0572	0.0196	0.0342	0.0508
			0.0303	0.1257	0.0546	0.0692	0.078	0.0574	0.0482	0.0449	0.0575	0.0909	0.0487	0.0672	0.0792
H03	0.021835	0.014436		0.0785	0.0356	0.0461	0.046	0.032	0.0243	0.0182	0.0231	0.0555	0.0232	0.0462	0.0564
				0.1129	0.0568	0.0729	0.0696	0.0564	0.0509	0.0446	0.0458	0.0791	0.0487	0.077	0.0848
H04	0.097572	0.110185	0.096952		0.0564	0.0737	0.0875	0.0721	0.0879	0.0747	0.0781	0.1357	0.0833	0.1091	0.102
					0.0837	0.102	0.1227	0.0985	0.1156	0.1053	0.1111	0.165	0.115	0.1404	0.1389
M01	0.037566	0.041702	0.046338	0.069733		0.0152	0.0208	0.035	0.0352	0.0335	0.0338	0.061	0.0226	0.0576	0.0627
						0.04	0.0421	0.0574	0.0573	0.059	0.0539	0.0862	0.0451	0.0834	0.0868
M02	0.052783	0.053349	0.060791	0.088507	0.026741		0.0342	0.0474	0.0452	0.039	0.0404	0.0808	0.0339	0.0661	0.0698
							0.0589	0.0728	0.0738	0.0662	0.066	0.1055	0.0573	0.0963	0.1011
M03	0.060324	0.065771	0.057456	0.10435	0.032014	0.044413		0.0596	0.0524	0.0464	0.0518	0.0851	0.052	0.076	0.0799
								0.0834	0.0763	0.0724	0.0737	0.1109	0.0769	0.1024	0.1077
NA10	0.042852	0.045232	0.044158	0.085815	0.046178	0.059396	0.071022		0.0162	0.0101	0.0132	0.0634	0.0111	0.046	0.0536
									0.0372	0.0339	0.0339	0.0882	0.0342	0.0741	0.0823
NA11	0.035397	0.033595	0.037838	0.101545	0.045572	0.060431	0.064487	0.027043		-0.0042	0	0.0541	0.0099	0.0432	0.0584
										0.0214	0.021	0.0784	0.0328	0.0704	0.0837
R01	0.029085	0.028951	0.032769	0.091066	0.044176	0.052309	0.05839	0.023077	0.009515		-0.0063	0.0363	0.0007	0.0405	0.0504
											0.0188	0.0603	0.0245	0.0664	0.0811
R02	0.035595	0.038946	0.036323	0.092721	0.044311	0.053617	0.062078	0.023741	0.009219	0.007332		0.0504	0.0061	0.0356	0.0549
												0.0791	0.0267	0.0602	0.0814
S	0.069199	0.07718	0.068393	0.149887	0.074425	0.090899	0.096628	0.076538	0.066531	0.050139	0.064363		0.0477	0.0777	0.0976
													0.0727	0.1026	0.1255
SP01	0.034442	0.035086	0.036864	0.099321	0.032147	0.046152	0.062937	0.021329	0.022566	0.012391	0.016705	0.060107		0.0357	0.0612
														0.0586	0.0872
SP02	0.059631	0.052941	0.063156	0.124758	0.071081	0.080385	0.089031	0.060803	0.056857	0.05496	0.047953	0.091057	0.048065		0.0627
															0.0908
SP03	0.082207	0.065793	0.073205	0.119818	0.072819	0.087055	0.09229	0.068396	0.07169	0.066331	0.066402	0.109925	0.072308	0.077956	

C.

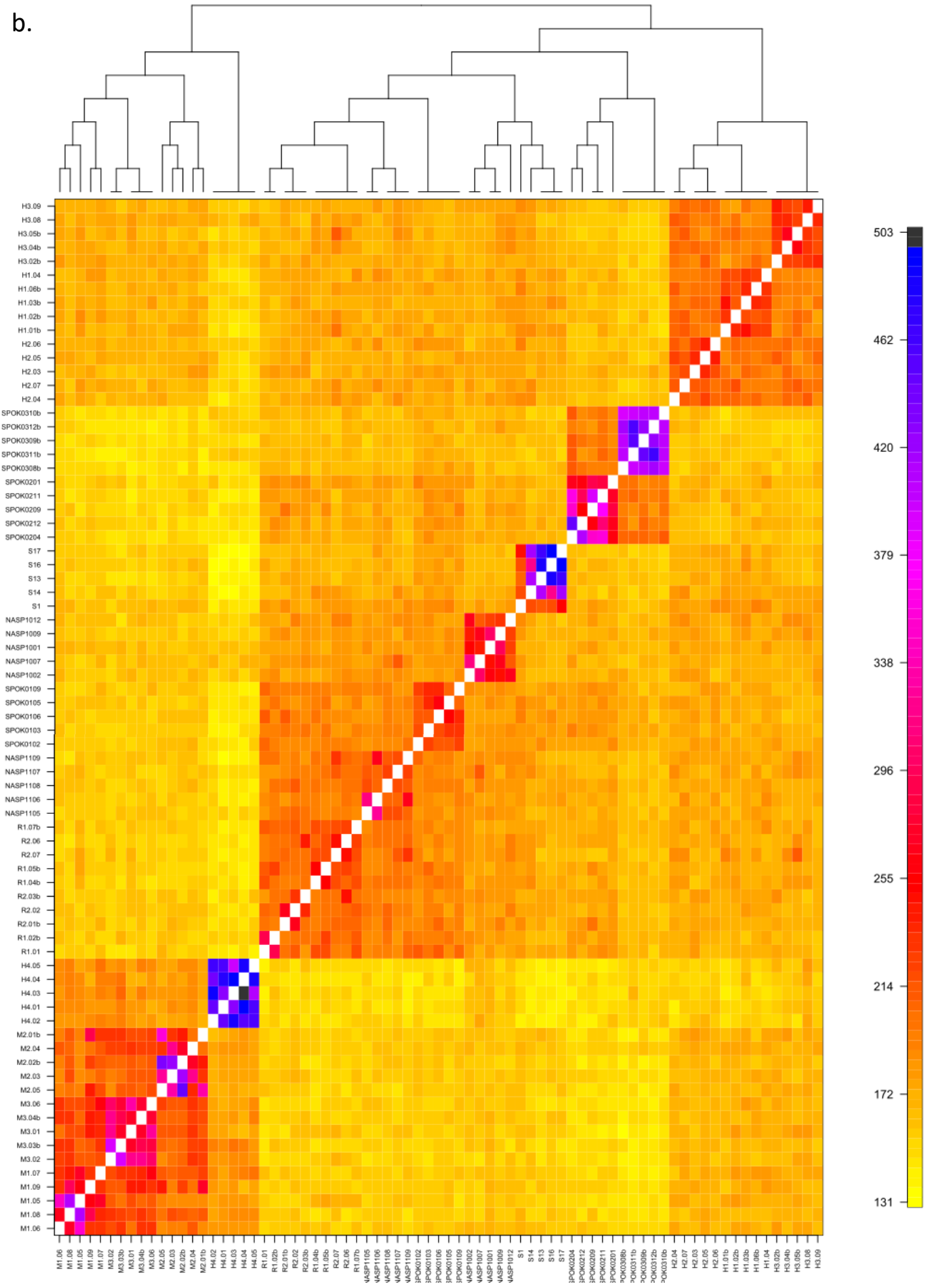
MAF10	H01	H02	H03	H04	M01	M02	M03	NA10	NA11	R01	R02	S	SP01	SP02	SP03
H01		0.0075	0.0057	0.0869	0.0308	0.0428	0.051	0.0284	0.0239	0.0169	0.029	0.0516	0.0236	0.0431	0.0771
		0.0431	0.033	0.1245	0.057	0.0718	0.0813	0.0586	0.0503	0.0453	0.0553	0.0794	0.0491	0.0837	0.1072
H02	0.025907		0.0033	0.1022	0.0283	0.0369	0.0556	0.0298	0.0152	0.0107	0.0259	0.049	0.017	0.0332	0.0534
			0.0359	0.1414	0.0543	0.0706	0.0873	0.0606	0.0465	0.0434	0.0561	0.0882	0.0502	0.0726	0.0855
H03	0.019584	0.018335		0.0829	0.039	0.0482	0.0496	0.0324	0.0286	0.0236	0.0295	0.0475	0.0286	0.0487	0.0638
				0.1205	0.0628	0.0737	0.0831	0.0644	0.0591	0.0545	0.057	0.0764	0.0607	0.0782	0.0928
H04	0.106301	0.122602	0.101731		0.0577	0.0763	0.093	0.0703	0.0881	0.0793	0.0855	0.1349	0.0908	0.1118	0.1047
					0.0908	0.1121	0.1304	0.1089	0.1216	0.1117	0.1199	0.1708	0.1247	0.1535	0.1405
M01	0.044716	0.04183	0.051571	0.073699		0.0178	0.0235	0.0388	0.0363	0.0314	0.0386	0.0588	0.0207	0.0613	0.0612
						0.0421	0.0451	0.0639	0.0619	0.0554	0.0606	0.0857	0.044	0.0949	0.0848
M02	0.057017	0.05405	0.062179	0.096026	0.030519		0.0397	0.0505	0.0498	0.0367	0.0462	0.0723	0.0335	0.0691	0.0819
							0.0672	0.0754	0.0747	0.0678	0.0723	0.1029	0.0585	0.1002	0.1092
M03	0.065882	0.067938	0.06397	0.111854	0.033889	0.051721		0.0574	0.0529	0.044	0.055	0.0845	0.0516	0.0783	0.0802
								0.0887	0.0776	0.0724	0.0804	0.1121	0.0789	0.1129	0.1089
NA10	0.042999	0.042807	0.045996	0.088925	0.049355	0.063186	0.071964		0.016	0.0105	0.0186	0.0588	0.0106	0.041	0.0561
									0.043	0.0355	0.0471	0.0922	0.0375	0.0754	0.0858
NA11	0.03726	0.030645	0.045067	0.105192	0.049143	0.061159	0.064707	0.028689		-0.005	0.0028	0.0545	0.0101	0.0431	0.0604
										0.0192	0.0243	0.0826	0.0331	0.0707	0.0918
R01	0.029432	0.026277	0.038362	0.096106	0.044467	0.051599	0.055572	0.02102	0.007451		-0.67	0.0395	-0.0032	0.0402	0.0548
											0.015	0.0699	0.0244	0.0734	0.0872
R02	0.041139	0.039937	0.041662	0.100717	0.049245	0.057782	0.066904	0.029128	0.014299	0.004693		0.0544	0.0063	0.0385	0.0587
												0.0781	0.0275	0.0676	0.0828
S	0.065084	0.070148	0.063462	0.155296	0.073006	0.08857	0.098562	0.076286	0.06734	0.052537	0.064774		0.0422	0.0748	0.0982
													0.073	0.1073	0.1302
SP01	0.035057	0.033627	0.042717	0.108001	0.032939	0.045012	0.065733	0.02277	0.022347	0.012491	0.015932	0.057926		0.0315	0.0599
														0.0598	0.0943
SP02	0.063505	0.053768	0.065071	0.134678	0.078402	0.087007	0.094794	0.059926	0.057557	0.056823	0.051812	0.091518	0.047127		0.0647
															0.0986
SP03	0.091488	0.067318	0.079575	0.124851	0.073426	0.0957	0.092522	0.070971	0.074407	0.069594	0.069003	0.112516	0.079186	0.080832	

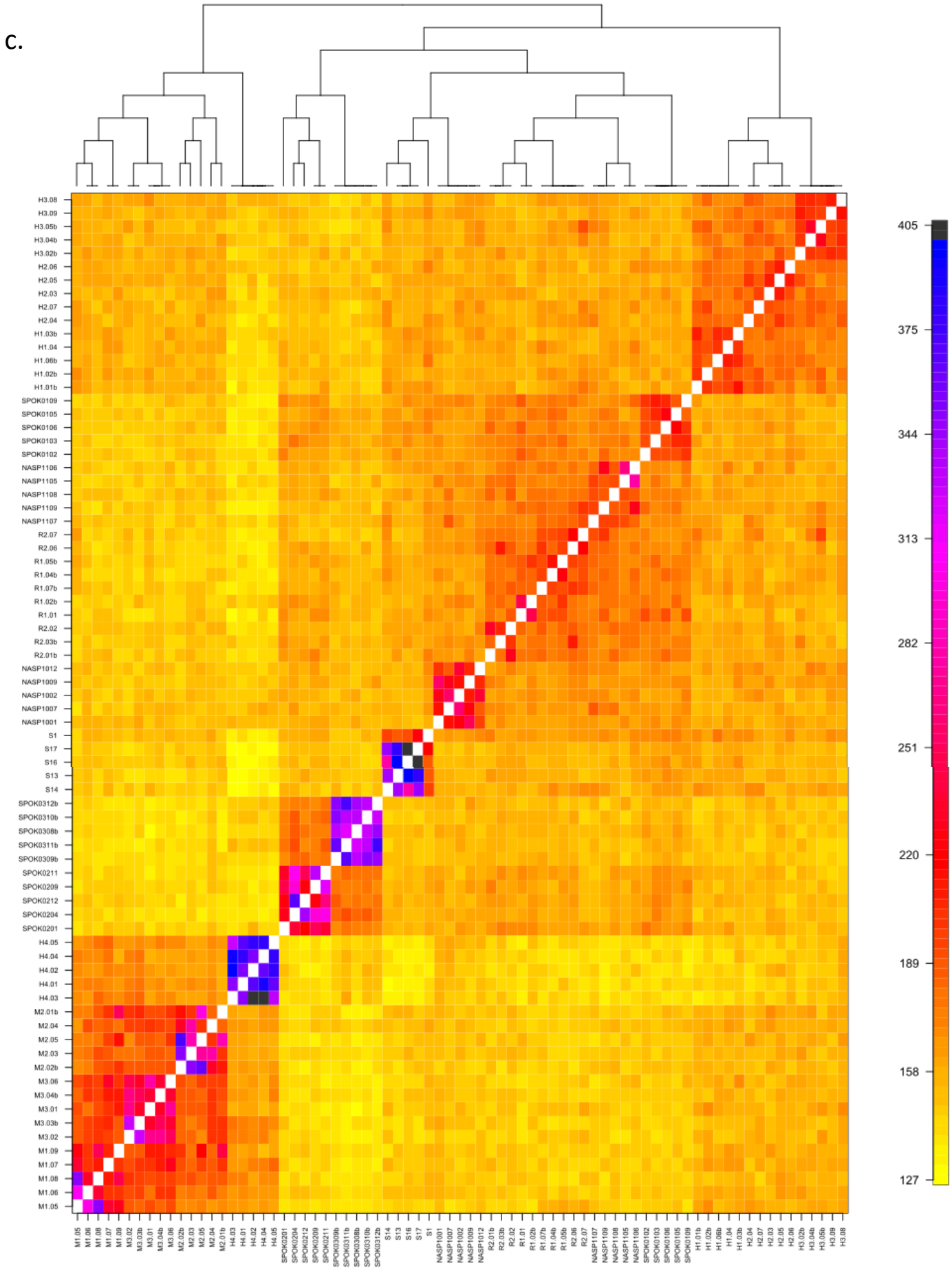
Supplementary Figure 4. Matrix of pairwise F_{ST}, mean F_{ST} values are presented in lower left section and confidence intervals (2.5%, 97.5%) in the upper right section. a) - c) are the same dataset but filtered with different MAF values (a) 0, (b) 0.05, (c) 0.1.

a.


SPOK0312b
SPOK0311b
SPOK0310b
SPOK0309b
SPOK0308b
SPOK0201
SPOK0211
SPOK0209
SPOK0204
SPOK0212
H3.08
H3.05b
H3.09
H3.04b
H3.02b
H1.06b
H1.02b
H1.03b
H1.04
H1.01b
H2.06
H2.05
H2.03
H2.07
H2.04
NASP1002
NASP1009
NASP1001
NASP1012
NASP1007
S17
S13
S14
S16
S1
SPOK0106
SPOK0105
SPOK0109
SPOK0103
SPOK0102
NASP1108
NASP1109
NASP1107
NASP1106
NASP1105
R2.07
R2.06
R1.07b
R2.03b
R2.01b
R2.02
R1.05b
R1.04b
R1.02b
R1.01
H4.02
H4.05
H4.03
H4.04
H4.01
M2.05
M2.03
M2.02b
M2.01b
M2.04
M3.06
M3.04b
M3.01
M3.03b
M3.02
M1.07
M1.09
M1.08
M1.06
M1.05
681
621
561
500
440
380
320
260
200
140

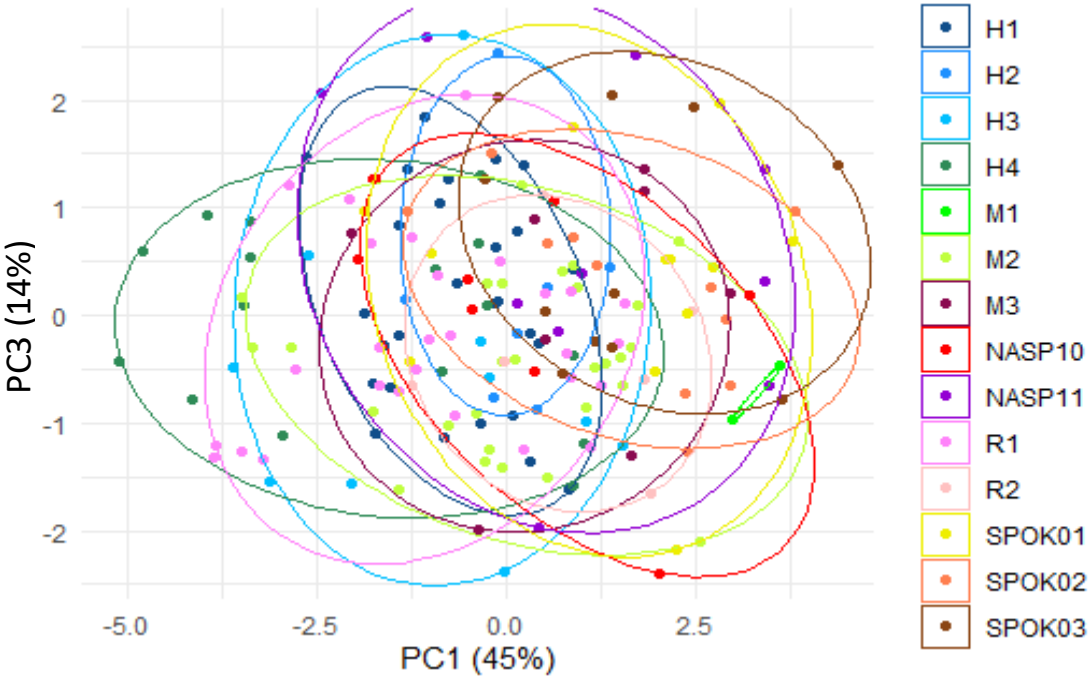
b.


H3.09
H3.08
H3.05b
H3.04b
H3.02b
H1.04
H1.06b
H1.03b
H1.02b
H1.01b
H2.06
H2.05
H2.03
H2.07
H2.04
SPOK0310b
SPOK0312b
SPOK0309b
SPOK0311b
SPOK0308b
SPOK0201
SPOK0211
SPOK0209
SPOK0212
SPOK0204
S17
S16
S13
S14
S1
NASP1012
NASP1009
NASP1001
NASP1007
NASP1002
SPOK0109
SPOK0105
SPOK0106
SPOK0103
SPOK0102
NASP1109
NASP1107
NASP1108
NASP1106
NASP1105
R1.07b
R2.06
R2.07
R1.05b
R1.04b
R2.03b
R2.02
R2.01b
R1.02b
R1.01
H4.05
H4.04
H4.03
H4.01
H4.02
M2.01b
M2.04
M2.02b
M2.03
M2.05
M3.06
M3.04b
M3.01
M3.03b
M3.02
M1.07
M1.09
M1.05
M1.08
M1.06
503
462
420
379
338
296
255
214
172
131

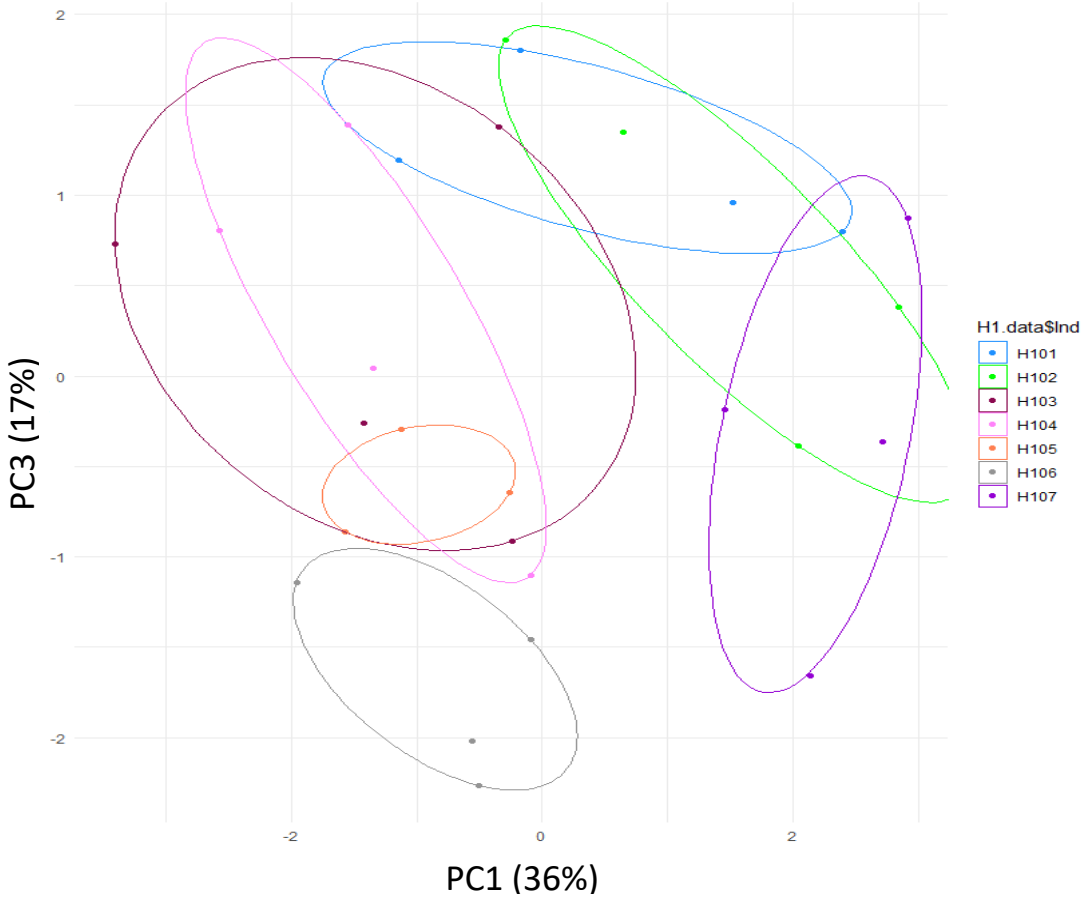


Supplementary Figure 5. Clustered fineRADstructure coancestry matrix. Individuals are listed along with the population from which they derive. The colour of each square represents the results of a pairwise comparison of estimated coancestry between two individuals based on similarity between RAD loci. The relative coancestry is illustrated, with high levels indicated by blue colouration and lower levels indicated by yellow. The phylogeny is illustrative of relationships between populations but does not represent true population history. a) - c) are the same dataset but filtered with different MAF values (a) 0, (b) 0.05, (c) 0.1.

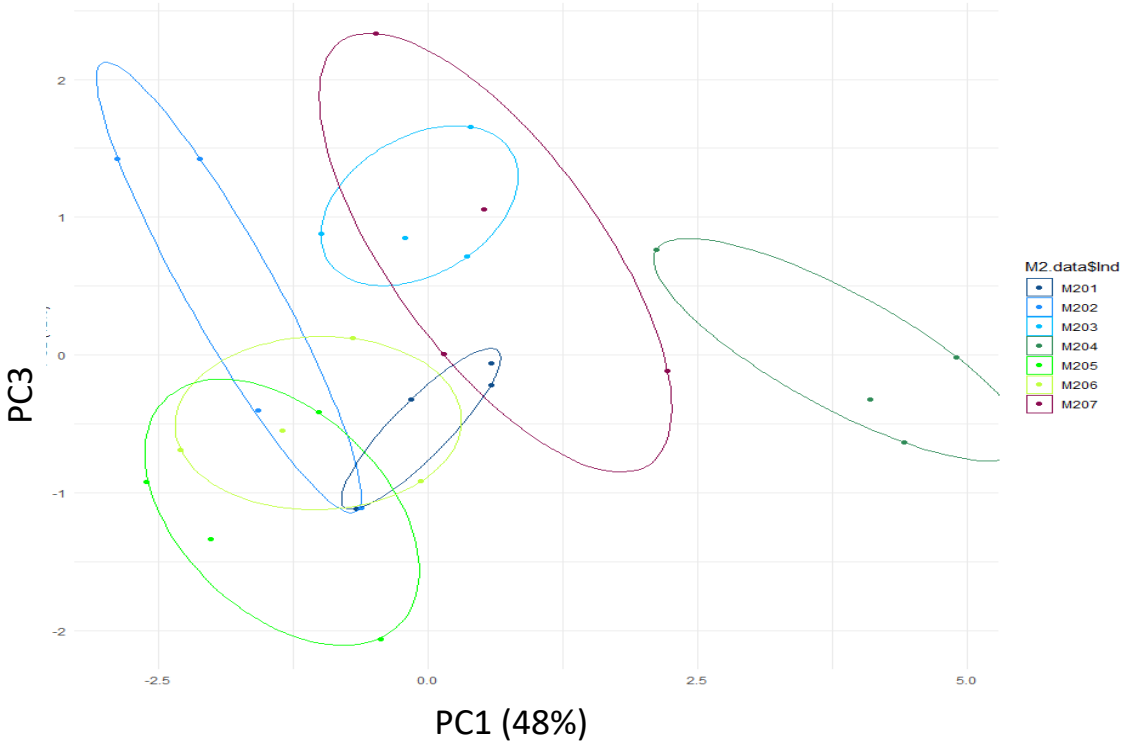
a.


PC3 (14%)
PC1 (45%)
2
1
0
-1
-2
-5.0
-2.5
0.0
2.5
H1
H2
H3
H4
M1
M2
M3
NASP10
NASP11
R1
R2
SPOK01
SPOK02
SPOK03

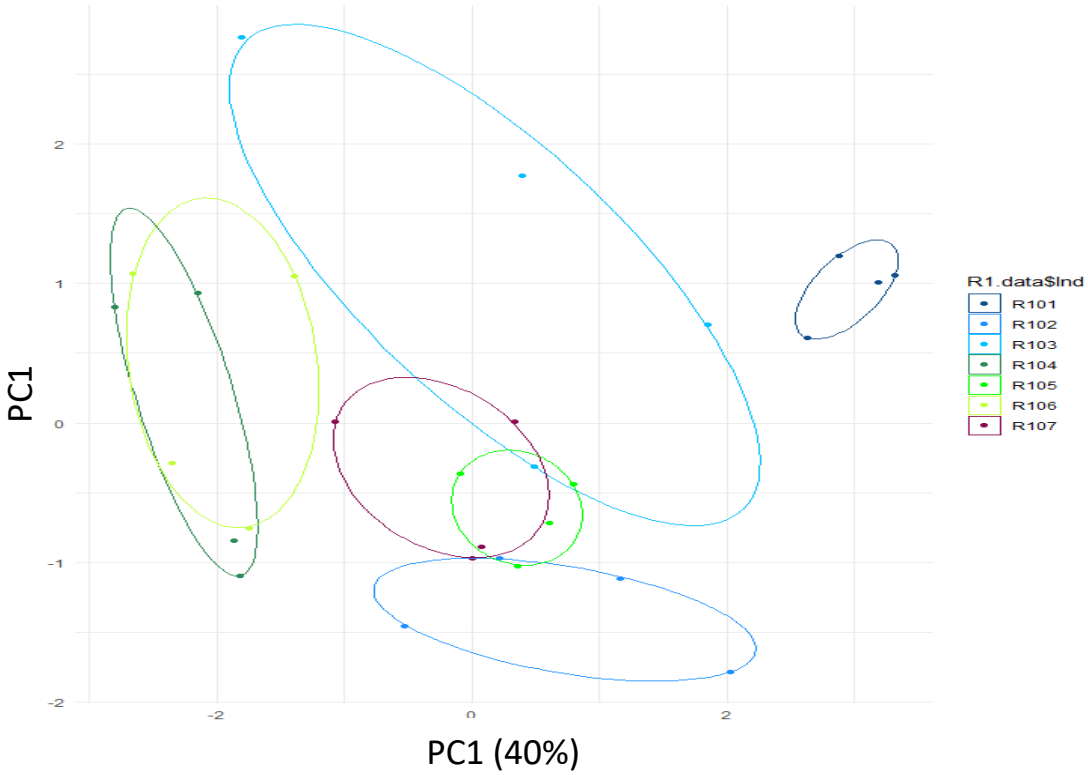
b.


PC3 (17%)
PC1 (36%)
2
1
0
-1
-2
-2
0
2
H1.data$Ind
H101
H102
H103
H104
H105
H106
H107

c.



d.



Supplementary Figure 6. Scatterplots illustrating variation in ray floret measurements. Principal component (PC) scores are along each axis, with values computed from a principal component analysis (PCA). Percentages indicate the amount of variation each PC explains. a) Each datapoint represents an individual and datapoints are colour coded by population. b) - d) Each datapoint represents a capitulum and datapoints are colour coded by individual. Every scatterplot b) - d) represents an individual population b) H01 c) M02 c) R01.

Pop	1	2	3	4	5	6	7	8
H1	0.593 ± 0.021	0.734 ± 0.019	3.187 ± 0.084	3.184 ± 0.082	3.545 ± 0.076	0.438 ± 0.016	0.450 ± 0.015	0.420 ± 0.013
H2	0.613 ± 0.024	0.684 ± 0.013	3.041 ± 0.074	3.031 ± 0.072	3.379 ± 0.120	0.468 ± 0.014	0.467 ± 0.016	0.448 ± 0.007
H3	0.604 ± 0.016	0.776 ± 0.029	3.303 ± 0.203	3.286 ± 0.201	3.474 ± 0.173	0.427 ± 0.026	0.429 ± 0.024	0.445 ± 0.023
H4	0.599 ± 0.026	0.786 ± 0.039	3.332 ± 0.112	3.325 ± 0.114	3.725 ± 0.155	0.422 ± 0.015	0.426 ± 0.014	0.412 ± 0.017
M1	0.671 ± 0.030	0.701 ± 0.006	2.320 ± 0.075	2.230 ± 0.083	2.682 ± 0.044	0.483 ± 0.034	0.526 ± 0.0135	0.511 ± 0.044
M2	0.628 ± 0.011	0.749 ± 0.017	3.158 ± 0.216	3.143 ± 0.210	3.287 ± 0.070	0.454 ± 0.021	0.452 ± 0.016	0.456 ± 0.010
M3	0.603 ± 0.042	0.729 ± 0.032	3.034 ± 0.077	3.010 ± 0.073	3.210 ± 0.102	0.470 ± 0.017	0.457 ± 0.019	0.473 ± 0.017
NA10	0.595 ± 0.023	0.720 ± 0.017	3.129 ± 0.125	3.124 ± 0.125	3.396 ± 0.130	0.483 ± 0.020	0.485 ± 0.020	0.453 ± 0.018
NA11	0.619 ± 0.019	0.662 ± 0.020	2.974 ± 0.114	2.965 ± 0.113	3.272 ± 0.130	0.517 ± 0.015	0.517 ± 0.016	0.468 ± 0.019
R1	0.546 ± 0.011	0.736 ± 0.035	3.138 ± 0.135	3.138 ± 0.133	3.362 ± 0.103	0.445 ± 0.024	0.454 ± 0.024	0.430 ± 0.011
R2	0.593 ± 0.018	0.729 ± 0.023	2.865 ± 0.141	2.871 ± 0.145	3.079 ± 0.145	0.493 ± 0.020	0.501 ± 0.022	0.463 ± 0.021
SP01	0.617 ± 0.017	0.710 ± 0.019	2.889 ± 0.094	2.880 ± 0.094	3.202 ± 0.111	0.512 ± 0.016	0.587 ± 0.016	0.483 ± 0.013
SP02	0.630 ± 0.015	0.673 ± 0.013	2.852 ± 0.084	2.846 ± 0.084	3.022 ± 0.103	0.492 ± 0.013	0.497 ± 0.014	0.491 ± 0.014
SP03	0.634 ± 0.017	0.697 ± 0.013	2.857 ± 0.093	2.851 ± 0.093	3.153 ± 0.088	0.520 ± 0.015	0.531 ± 0.016	0.468 ± 0.011

Supplementary Table 2. Mean floral trait values per population ± S.E. These traits were used in the principal components analysis presented in Fig 3.6. Traits are as follows: (1) Ratio between the height of the spot and the spotted ray floret, (2) Ratio between the height of the spotted ray and the plain ray floret, (3) Aspect ratio of the spot, (4) Aspect ratio of the spotted ray floret, (5) Aspect ratio of the plain ray floret (6) Circularity of the spot, (7) Circularity of the spotted ray floret, (8) Circularity of the plain ray floret.

Appendix 4. Chapter 4 Supplementary Information

Trait	PC1	PC2	Cal	Spring
1	-0.34965	0.154689	1	0.18 ± 0.005
2	0.287551	0.401859	209 ± 1.52	226 ± 0.98
3	0.352264	0.128093	72.8 ± 1.1	119 ± 2.09
4	-0.1808	-0.38683	0.32 ± 0.006	0.26 ± 0.007
5	-0.1007	0.426266	20 ± 1.09	12.8 ± 1.09
6	0.336673	0.34084	101 ± 0.93	121 ± 1.02
7	0.328455	-0.1551	1.63 ± 0.032	2.96 ± 0.043
8	0.299936	-0.2561	1.37 ± 0.027	1.87 ± 0.03
9	-0.28482	0.374385	0.51 ± 0.005	0.37 ± 0.008
10	0.332954	-0.27192	1	0.55 ± 0.01

Supplementary Table 3. The relative loadings of floral trait measurements from the Cal and Spring morphotypes on each principal component (PC1 and PC2) for the analysis presented in Fig 4.6. Traits are as follows (1) Proportion of ray florets within a capitulum that are spotted. (2) Red values (from red-green-blue extracted values) of the plain segment of the spotted ray floret. (3) Green values (from red-green-blue extracted values) of the plain segment of the spotted ray floret. (4) The length of the plain segment of the spotted ray floret (from the top of the spot to the tip of the ray floret). (5) Blue values (from red-green-blue extracted values) of the plain segment of the spotted ray floret. (6) Mean brightness value on the plain segment of the spotted ray floret. (7) Aspect ratio of the spotted ray floret. (8) Aspect ratio of the spot. (9) The length of the plain segment of the spotted ray floret as a proportion of the total ray floret length. (10) The area of the spot as a proportion of the total spotted ray floret area. The mean ± S.E. trait values are given for Cal and Spring morphotypes.

Appendix 5. Chapter 5 Supplementary Information

Supplementary Figure 7. T_1 *Nicotiana tabacum* stably transformed with *GdMYB8* genes on a constitutive promoter (35S::*GdMYB8*). a) Representative leaves and flowers from one line of *GdMYB8a*, *GdMYB8b*, *GdMYB8c* transformants compared to wild type. The line presented here was chosen as typical for each set of transformants, the full range of phenotypes across lines is presented in the main text.

Alignment Position	Amino Acid
96	T/I
127	H/Y
133	H/Y/N
148	P/S
150	Y/S
152	T/S
182	A/S
406	W/G
190	A/V
267	M/R
269	K/Q
286	L/I
314	N/D
354	A/D
356	S/L
392	S/A

Supplementary Table 4. Positions in GdMAT1 alignment (position one being the M encoded by the start codon) were there was variation in the amino acid sequence present.

Alignment Position	Amino Acid
25	R/STOP
44	E/K
45	K/N
50	V/A
72	F/S
85	E/V
109	L/P
113	R/W
138	H/N
144	D/E
151	L/STOP
152	D/E
214	I/V
269	E/K
274	E/G
288	K/T

Supplementary Table 5. Positions in GdDFR alignment (position one being the M encoded by the start codon) were there was variation in the amino acid sequence present. STOP indicates the position of the stop codons in the gDNA variants thought to be pseudogenes.

Transgenic *N. tabacum* gene expression data t-tests

Results of the t-tests testing whether the expression levels determined through qRT-PCR, of *NtANS*, *NtDFR*, and *NtMAT* in transgenic tobacco (transformed with either *GdMYB8a, GdMYB8b, GdMYB8c* on a constitutive promoter), differ significantly from wild type *N. tabacum* expression levels in petal tissue. All pairwise comparisons are with wild type *N. tabacum*: *NtANS* (*GdMYB8a* t=2.37 p=0.03, *GdMYB8b* t=7.77 p<0.0001, *GdMYB8c* t=5.13 p=0.0002), *NtDFR* (*GdMYB8a* t=5.39 p=0.0001, *GdMYB8b* t=8.3 p<0.0001, *GdMYB8c* t=5.96 p=0.0001), *NtMAT1* (*GdMYB8a* t=-0.60 p=0.66, *GdMYB8b* t=-1.73 p=0.30, *GdMYB8c* t=-1.95 p=0.30).

Appendix 6. Chapter 6 Supplementary information

G. diffusa stable transformation media

cSIM

For 1L: 4.4g Murashige-Skoog Medium with vitamins (Duchefa), 30g sucrose, 0.5mg/ml silver nitrate (AgNO3), 20mg/ml ascorbic acid, 0.5g MES, 30mg L-cysteine, 25mg/l kanamycin, 1mg/ml IAA, 2mg/ml *trans*-Zeatin riboside, 0.5mg/ml 2, 4-Dichlorophenoxyacetic, pH 5.9.

rSIM

for 1L: 4.4g Murashige-Skoog Medium with vitamins (Duchefa), 20g sucrose, 20mg/l ascorbic acid, 30mg/l L-cysteine, 25mg/l kanamycin, 250mg/l cefotaxime, 1mg/ml IAA, 2mg/l *trans*-Zeatin riboside, pH 5.9

SIM

for 1L: 4.4g Murashige-Skoog Medium with vitamins (Duchefa), 20g sucrose, 20mg/l ascorbic acid, 30mg/l L-cysteine, pH 5.9

MS9

for 1L: 4.4g Murashige-Skoog Medium with vitamins (Duchefa), 20g sucrose, 0.5mg/l IAA and 1mg/l BAP

CCM

for 1L: 4.4g Murashige-Skoog Medium with vitamins (Duchefa), 30g sucrose, 2mg/l IAA, 0.25mg/l *trans*-Zeatin riboside, 0.5mg/l 2, 4-Dichlorophenoxyacetic, pH 5.9

Yeast one-hybrid media

YPDA

for 1L: 20g peptone, 10g yeast extract, 75mg adenine, 2% (v/v) glucose, 100ml amino acid solution (Appendix 6), 20g agar

Gel-shift buffers

Laemmli buffer

for 1L: 144.2g glycine, 30.3g tris base, 0.5% (w/v) SDS, pH 8.3

Coomassie blue

for 1L: 100ml glacial acetic acid, 450ml ddH_2O, 3g Coomassie Dye (Brilliant Blue G, Sigma), 450ml methanol

Base buffer

10mM Tris pH8, 250mM NaCl, 5% (v/v) glycerol

Lysis buffer

for 25ml: 24ml base buffer, 50µl 1M DTT, 1% (w/v) Sarkosyl, 250µl 100mM PMSF

Loading buffer
for 25ml: 24ml base buffer, 50µl 1M DTT, 1% (w/v) sarkosyl, 8.7mg imidazole

Washing buffer
for 50ml: 49ml base buffer, 100µl 1M DTT, 0.25g sarkosyl, 109mg imidazole

Annealing buffer
100mM tris pH7.5, 1.5M NaCl, 10mM EDTA pH8

Binding buffer
for 50ml: 1ml 1M tris pH8, 1.5ml 5M NaCl, 25µl 0.5M EDTA pH8, 100µl 1M $MgCl_2$, igepal (NP-40) 10µl, 1% (v/v) glycerol, ddH_2O up to 50ml

Acrylamide gel
for 12ml: 1.8ml acrylamide (29:1), 600µl 10x TBE, 120µl 10% (v/v) APS, 12µl TEMED, 9.6ml ddH_2O

Gel-shift acrylamide gel buffers tested

1) (for 50ml) Tris 500ul (10mM), KCl 2.5ml (50mM), DTT 50ul (1mM), ddH_2O 46.95ml
2) (for 50ml) Tris pH8 1ml (20mM), NaCl 50ul (10mM), EDTA 200ul (2mM), DTT 100ul (2mM), glycerol 5ml (10% v/v), ddH_2O 43.65ml
3) (for 10ml) KCl 750ul (150mM), DTT 0.5ul (0.1mM), EDTA 1ul (0.1mM), Tris 100ul (10mM), ddH_2O 4.2ml
4) (for 50ml) Tris 500ul (10mM), NaCl 250ul (50mM), DTT 50ul (1mM), EDTA 100ul (1mM), glycerol 2.5ml (5% v/v), ddH_2O 46.6ml

Y1H 10x dropout solution
In -HIS media '1' is removed, in -HIS -LEU '1' and '2' are removed.

Amino acid	**10x Concentration**
L-Adenine hemisulfate salt	200mg/L
L-Arginine HCl	200mg/L
L-Histidine HCl monohydrate[1]	200mg/L
L-Isoleucine	300mg/L
L-Leucine[2]	1000mg/L
L-Lysine HCl	300mg/L
L-Methionine	200mg/L
L-Phenylalanine	500mg/L
L-Threonine	2000mg/L
L-Tryptophan	200mg/L
L-Tyrosine	300mg/L
L-Uracil	200mg/L
L-Valine	1500mg/L

Solvent Gradient used in HPLC

Time/mins	Solvent A % (0.5% formic acid)	Solvent B % (Acetonitrile)
0	95	5
2	95	5
42	0	100
47	0	100
48	95	5
53	95	5