INVESTIGATING THE MADS-BOX PROTEIN FAMILY IN THE FERN *CERATOPTERIS RICHARDII*.

By

Derry Carr

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
September 2024

DECLARATIONS

ABSTRACT

The family of transcription factors known as MADS-box proteins is found across all land plants, with multiple MADS-box encoding genes per species. In flowering plants such as *Arabidopsis thaliana* a subset of MADS-box proteins regulates the development of flowers and seeds, binding to each other following the 'ABCDE' floral model to specify the identities of each organ type. These functions are thus important to food security because seeds are staple foods globally. MADS-box gene functions remain uncategorised in non-flowering plants, with no corresponding model to the 'ABCDE' found in flowering plants. How MADS-box floral functions evolved is largely unknown. This can be determined by studying MADS-box gene functions in non-flowering plants. The model fern *Ceratopteris richardii* falls into the sister clade of all seed-bearing plants and is an ideal candidate to investigate MADS-box genes in a non-flowering plant because gene functions can also now be tested directly in this species.

The phylogenetic data presented in this thesis suggests the MADS-box protein family present in *C. richardii* reflects the relationships observed by the MADS-box protein family of *A. thaliana*. The expression data shows differing expression between developmental stages, despite previous studies suggesting a ubiquitous expression of MADS-box proteins in non-seed plants. Protein-protein interactions within the MADS-box protein family in this thesis suggest no interaction contradicting the interactions observed in flowering plants and within the 'ABCDE' model. This thesis provides novel insights into the relationship, expression, and interactions of MADS-box proteins in *C. richardii*. The work in this thesis offers a starting point for future investigation into MADS-box proteins in non-seed plants.

ACKNOWLEDGMENTS

I would first and foremost like to thank my supervisor Dr Juliet Coates for her unwavering support and her ability to go above and beyond in the face of adversity to help me greatly throughout my PhD. Without her I truly would not have been able to present this thesis. Because of her I have developed not only as a researcher and a scientist but as a person and I will forever be grateful for everything she has done for me both academically and beyond.

I would also like to thank Professor Daniel Gibbs and the Gibbs lab members for their wonderful support as a lab neighbour. Dr Andrew Plackett for his *Ceratopteris* knowledge and materials. A huge thank you must also be given to the Central Services and Stores staff for their contribution to the day-to-day workings of the Plant Science theme. Finally, I would like to thank my fellow lab member Kirsty McCready for her support for my wellbeing as well as my work.

In addition, I would like to thank the University of Birmingham for funding my project, to the Genetics Society and the University of Birmingham for supporting my attendance of the NIBB Trans-Scale Biology using exotic non-model organisms EMBO workshop in Okazaki, which will be an experience I cherish forever.

And thank you so much to my friends and family for being endlessly supportive and patient as I navigated this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1. Introduction

1.1.1. Food security and the importance of seeds.

Around 80% of the world's nutrients are obtained from seed-bearing crops (United Nations, 2019). Therefore, it has been a priority to understand the genes and mechanisms behind how spermatophytes (seed plants) reproduce (Pereira and Coimbra, 2019). With climate change becoming a bigger issue in recent years, this understanding is more important than ever. The ability to genetically modify crops to be able to survive in more challenging conditions than previously is important in ensuring food security in the future (Qaim and Kouser, 2013). Stresses such as flooding (Gao *et al* 2019), drought and high temperature can impact plant reproductive development and hence seed production (Liu *et al* 2023) and lead to genetic trade-offs between stress resistance and reproduction (Hatfield and Prueger, 2015).

Agriculture has been the main source of food for humanity since the neolithic era. The "Neolithic Revolution" was the first farming revolution occurring around 12,000 years ago. Since then, humanity has relied on crop plants more and more as a nutrient source (Olsson and Paik, 2016). The increased availability of food had an impact on the overall global population leading to an increase that has only been matched by the industrial revolution's impact on population, which coincided with the second agricultural revolution around the 18th century (Kerridge, 2013; Milner and Boldsen, 2023). The third agricultural revolution, the "Green Revolution" of the 1960s saw the introduction of selectively bred crops with higher seed yields, with the main crops being corn, wheat, and rice, in different global locations (Pingali, 2012).

The Green Revolution has been attributed with reduction in poverty and famine and an increase in food security (John and Babu, 2021). Beginning in Mexico, it was soon encouraged and expanded to the rest of South America, India, and Asia (Pingali, 2012). Southern America gained support from the more industrial USA due to the USA wishing to counterbalance the “Red Revolution” during the cold war (Lorek, 2022). At the same time, Asian governments implemented policies to protect young agricultural businesses and support the progress to improve infrastructure (Frankema, 2014). Despite this, African countries without independence faced no government subsidies for the necessary equipment such as improved fertilisers and irrigation (Pingali, 2012). Without pre-existing infrastructure and with a harsher climate many African countries experienced losses in the battle against food scarcity (Elhassnaoui *et al*., 2023).

The Green Revolution was a response to the insecurity of food accessibility globally, however the problems facing current food security is less in the abundance and more in the losses due to the unstable climate (Raza *et al*., 2019). With the reliance humanity has placed time and again on agriculture crops, combatting the problems being faced due to increasing temperatures, increasing floods, and other climate emergencies must take priority.

1.1.2. Where did seeds come from?

Despite extensive knowledge of the processes responsible for seed production and dispersal (Selwyn, Pino and Espelta, 2023), how seed-bearing plants including flowering plants (Angiosperms) evolved is less well understood. The ancestral system of reproduction that uses spores for dispersal of plant spores has been researched via

the fossil record (Bowles *et al*., 2022) and using extant spore-bearing plants such as the representative bryophyte *Physcomitrium* (formerly *Physcomitrella*) (Vesty *et al* 2016). The switch between the two systems of reproduction remains elusive but likely involved the evolution of heterospory (Linkies *et al*., 2010). In this project, we use a model organism from the sister group to seed-bearing plants to attempt to decipher this mystery.

1.2. Plant evolution

Plants most likely evolved from unicellular algae (Morris *et al*., 2018) (Figure 1.1). When plants colonised the land ~470 MYA they needed to adapt to this new dry world they found themselves in (Rensing, 2018). As such their previous reliance on water to reproduce needed to change. In bryophytes, lycophytes and extant ferns, sperm cells are flagellated, and fertilisation requires water, but the formation of desiccation-resistant spores allowed dispersal of the next generation in the absence of water (Eichorn and Evert, 2013).

Around 390 MYA plants were becoming more recognisable by today's standards, with roots and leaves being present. By the late Devonian period approximately 370 MYA, vascular (transport) tissue had evolved with early seeds also appearing in the lineage of now-extinct seed ferns (Harrison and Morris, 2017; Linkies *et al* 2010). In the carboniferous period before trees as we know them had taken hold of the land, giant fern-like plants including seed ferns dominated the landscape the domination of these plants and others like them led to a decrease in atmospheric carbon and a surplus of oxygen being produced, altering the climate recognisable with the Cretaceous period,

and ultimately resulting in a global forest dieback after the Chicxulub impact (Vajda and McLoughlin, 2004; Wilson *et al*., 2020). Seed fern extinction has been suggested to be correlated not only with these changes to the environment, but also to have been brought about due to the competition from the newly diversifying flowering plants (McLoughlin *et al*., 2008). However, extant ferns reproduce via spores (Figure 1.9) and the earliest extant seed plants (gymnosperms) appeared approximately 380 MYA (Linkies *et al* 2010) (Figure 1.1).

In seed plants, early diverging lineages such as Ginkgo and cycads retain flagellated sperm but the evolution of heterospory and hence pollen enabled fertilisation without the need for water, for example via wind or insect pollination (Eichorn and Evert, 2013). Post-fertilisation, the evolution of the complex multicellular seed structure to house the embryo enabled desiccation resistance, dormancy, and dispersal over large distances (Finch-Savage and Leubner-Metzger 2006; Linkies *et al* 2010).

Flowering plants (Angiosperms) had a chance to dominate plant reproductive strategies from around 200 MYA, after the Permian-Triassic extinction event (Blomenkemper *et al*., 2018). From within the Angiosperms, the changes in climate and reduction in CO_2 levels allowed grasses (Poaceae) to evolve around 100 MYA. From within the grasses, which are the most economically important plant family, humans have domesticated cereal crops for seed harvest for food and animal feed, in particular wheat, rice, maize, barley, oats, sorghum and millet (Wang *et al*., 2015b). All of these evolutionary changes in plants can be tracked through fossil records as well as through research into the evolution and development of extant plants (Kellogg, 2001; Wu, You and Li, 2017).

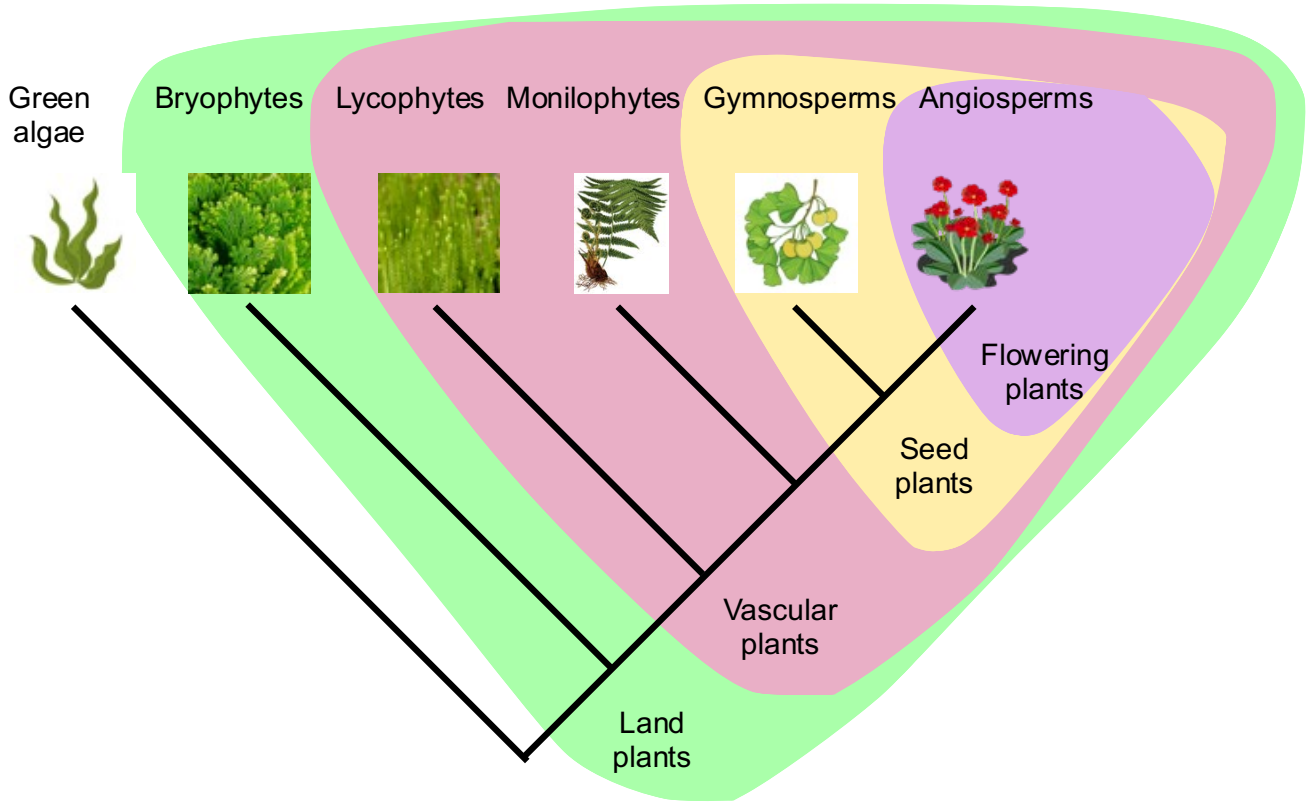


Figure 1.1: A schematic phylogenetic tree showing the evolution of plants. The green highlight covers all land plant groups, the pink vascular plants, the yellow seed plants, and the purple flowering plants.

The environment is currently going through another huge shift and humanity is reliant on extant plants. As such, an understanding of how plants adapted to previous climates and the application of this knowledge to the adaptation of extant plants, particularly crops, is necessary for humanity's survival.

1.3. Life cycles of flowering plants

After the colonisation of land, from about 500 MYA (Kapoor *et al*., 2022), some plants followed an evolutionary path that majorly reduced the gametophyte stage, transforming from a gametophyte-dominated life cycle (Figure 1.9) to a sporophyte-dominated life cycle (Figure 1.2). With this change, the evolution of the reproductive system could be observed, with the vastly reduced size and life-span of the gametophyte stage rendering it dependent on the sporophyte mother plant (Pereira and Coimbra, 2019). This, along with the evolution of heterospory, the production of differently sized female and male spores that give rise to dimorphic gametophytes (Wang and Bai, 2019; Leslie and Bonacorsi, 2022) enabled seed-, and then flower evolution. Heterosporous plants paved the way for the much more drastic size differences observed in flowering plants with the megaspore contained in the ovule (Figure 1.2) and the microspore on the stamen and in pollen (Pereira and Coimbra, 2019).

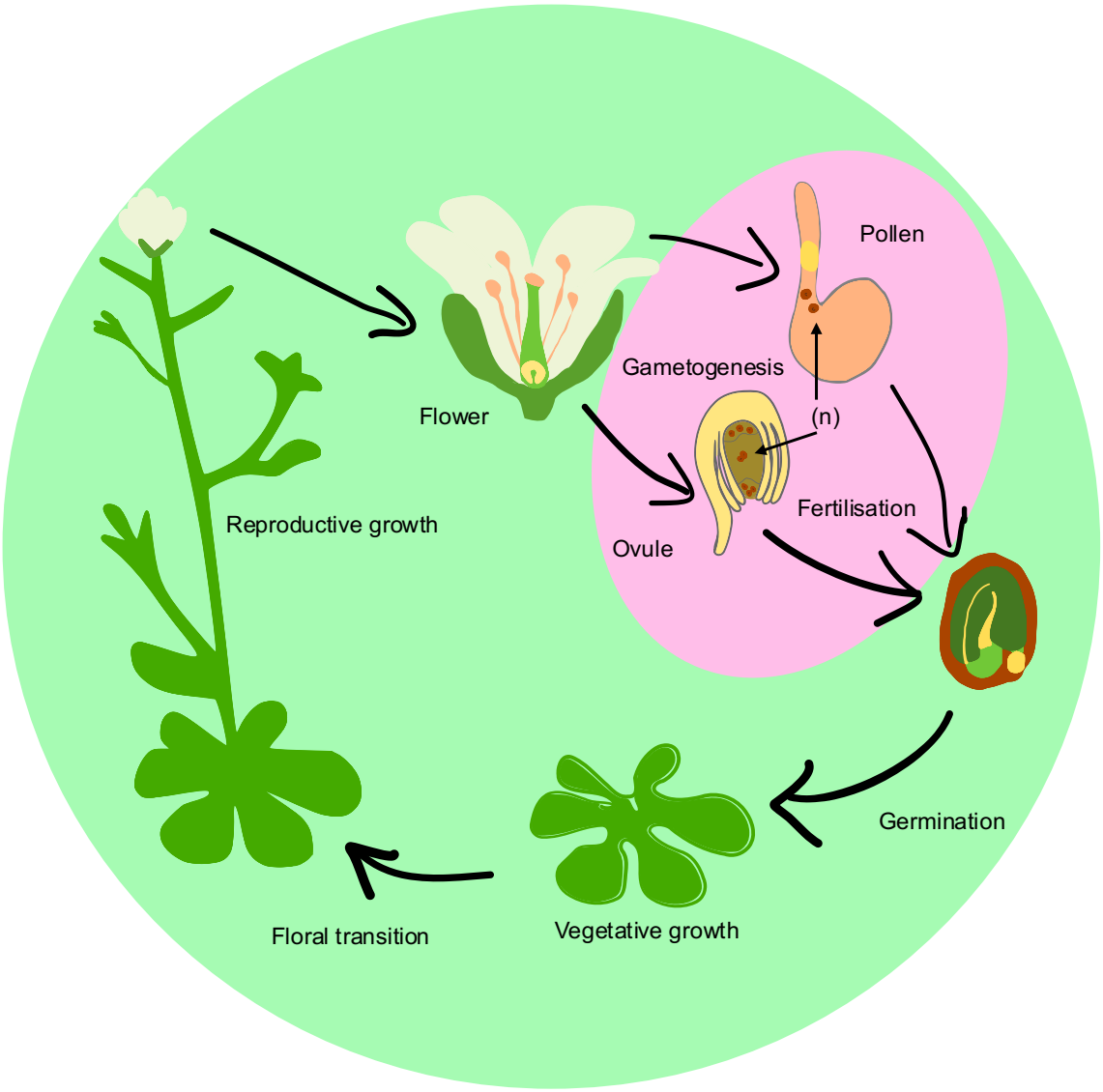


Figure 1.2: *Arabidopsis* life cycle as an example of a flowering plant life cycle highlighting the much-reduced gametophyte stages in the pink circle. The female gametophyte is the embryo sac, a 7-celled structure (brown) with 8 nuclei (red) housed inside the ovule, which also contains diploid sporophytic maternal tissue (yellow) surrounding the embryo sac. This diploid tissue later forms the seed coat. The male gametophyte consists of the pollen grain, which contains 2 haploid sperm cells for double fertilisation with the egg and with two polar nuclei from the embryo sac, to form the diploid embryo and the triploid endosperm respectively, and also contains the vegetative nucleus, which directs growth of the pollen tube.

1.4. Introduction to flowers

Flowers are the reproductive organ of angiosperms. Flowers develop from an apical meristem, but flower development is not initiated until the appropriate inductive signals for flowering, both environmental signals and endogenous signals such as plant hormones and sugars, are received by the plant (Vinicius and Schmid, 2014). At this point, the apical meristem converts from vegetative growth to reproductive growth. As such flowers are homologs of whole shoots, and individual floral organs are leaf homologs that have become specialised in sexual reproduction, (Specht and Bartlett, 2009). Flowers can be bisexual, containing both the female pistil and male stamens, or unisexual, (Specht and Bartlett, 2009). It is unknown whether bisexual or unisexual flowers evolved first with one theory suggesting unisexual flowers formed through a compression of secondary structures of a bisexual flower, whilst another suggests bisexuality evolved from a unisexual axis (Endress and Doyle, 2009). Both theories rely on a convergent model of evolution in the expression of B class genes linked to the 'ABCDE' model, (Zhang et al., 2022).

Another characteristic of flowers where the evolutionary origin is unknown is in the phyllotaxy. There are two distinct groups of flowers: whorled and spiralled (Wang *et al*., 2015a). These further break down into three basic patterns of floral phyllotaxis – simple whorl, complex whorl, and spiral (Endress and Doyle, 2015) (Figure 1.3). In whorled flowers, the organs are distinct and rely on different transcription factors in their tissue development. In spiral flowers such as those found in *Arabidopsis* the organs (sepals, petals, stamens, and carpels) are not equidistant from one another,

the transition between organs is often not as abrupt as in whorled flowers, and intermediate organs are relatively common, (Endress and Doyle, 2007).

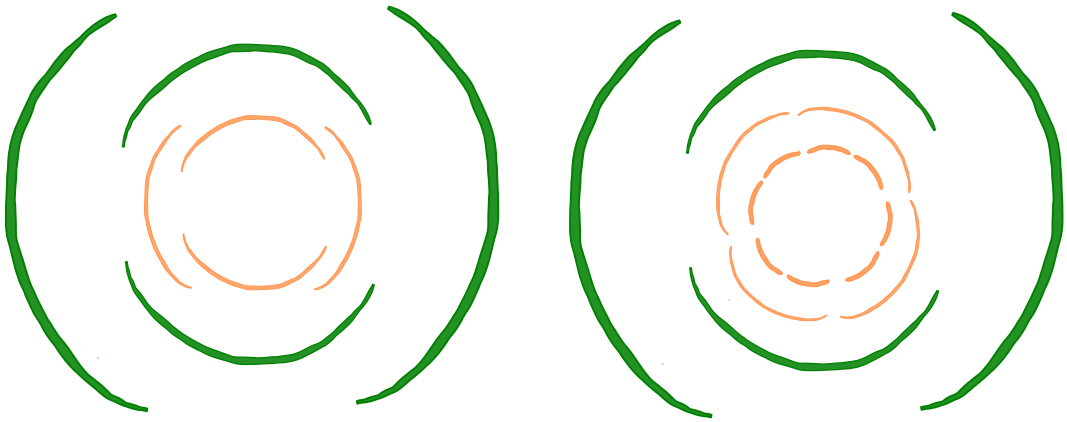


Figure 1.3: Two kinds of whorled phyllotaxis found in angiosperms The left shows simple whorls (those seen in *Arabidopsis*) whilst the right shows complex whorls (those seen in *Magnoliales*) with an increase in organ number per whorl. Green whorls represent the outer whorls sepals and petals, the orange whorls represent the inner whorls stamens and carpel. Adapted from "Ancestral traits and specializations in the flowers of the basal grade of living angiosperms" by Endress and Doyle (2015).

1.5. MADS-box genes and floral organ identity

The MADS-box family contains transcription factors that have been identified to be responsible for floral organ identity and development (Parenicová *et al.,* 2003). The ability to manipulate these genes to ensure crops become sexually mature even under environmental stressors is one way of increasing food security (Qaim and Kouser, 2013). The MADS-box gene family has been present since the most recent common ancestor (MRCA) of algae and land plants approximately 450 million years ago (MYA),

with functions in early land plants most likely being across early cell development and differentiation. It was not until the advent of flowers around 150 MYA (Chanderbali *et al*., 2016) that the genes gained more specialised functions, with different family members becoming responsible for not only the floral organ and fruit identity but also the timing of the development of these sexual organs (Lovisetto *et al*., 2011). However, despite the functions of this family being known in divergent land plants, namely bryophytes and angiosperms, there is a gap in knowledge in the plants that fall 'between' these two groups, in the sister clade to seed-plants, the monilophytes. Understanding the functions of this important gene family in the non-seed lineage closest to seed plants could provide understanding of floral and seed evolution at a deeper level allowing for more targeted genetic modification and breeding strategies in angiosperms in the face of ongoing climate change.

MADS-box genes have been categorised via the 'ABCDE' floral model to explain their involvement in the regulation of floral development. MADS-box transcription factors determine what organs shoot tissues will differentiate into (Figure 1.4). Extensive research on MADS-box gene roles in whorl formation and organ identity during flower development has been achieved in dicots such as *Arabidopsis*, *Anthirrhinum* (snapdragon) and the monocot and key crop plant wheat (*Triticum aestivum*) (Callens *et al*., 2018). The 'ABCDE' model categorises transcription factors into distinct groups depending on their function in specifying floral organs and different combinations of MADS-box genes are involved in the development and identity of the different whorls, namely sepals, petals, stamens, carpels, and ovules of flowering plants, (Murai, 2013) (Figure 1.4). Genetic approaches established the necessity of the E-class MADS-box genes *SEPALLATA* for floral organ identity and of the 'D'-class gene *SEEDSTICK* for

ovule specification, extending the original ‘ABC’ model (Coen and Meyerowitz 1991; Bowman and Moyroud 2024) (Figure 1.4). Biochemical assays established the higher-order complexes required for these interactions (Pelaz *et al*., 2000; Krizek and Fletcher, 2005).

Conservation of function of MADS-box genes is present between reproductive tissues of *Arabidopsis* and rice, suggesting MADS-box genes play an integral role in plant reproduction (Gramzow and Theißen, 2010). In addition to their functions in floral development, MADS-box genes also have a range of functions from cell development to auxin metabolism, discovered in the model plant *Arabidopsis* (Zhao *et al*., 2021). Conservation of function of MADS-box genes is present between reproductive tissues of *Arabidopsis* and rice, suggesting MADS-box genes play an integral role in plant reproduction (Gramzow and Theißen, 2010). Outside of sexual reproduction, MADS-box genes regulate development, growth, and function in flowering time, as well as floral organ identity (Theißen, Melzer, and Rümpler, 2016).

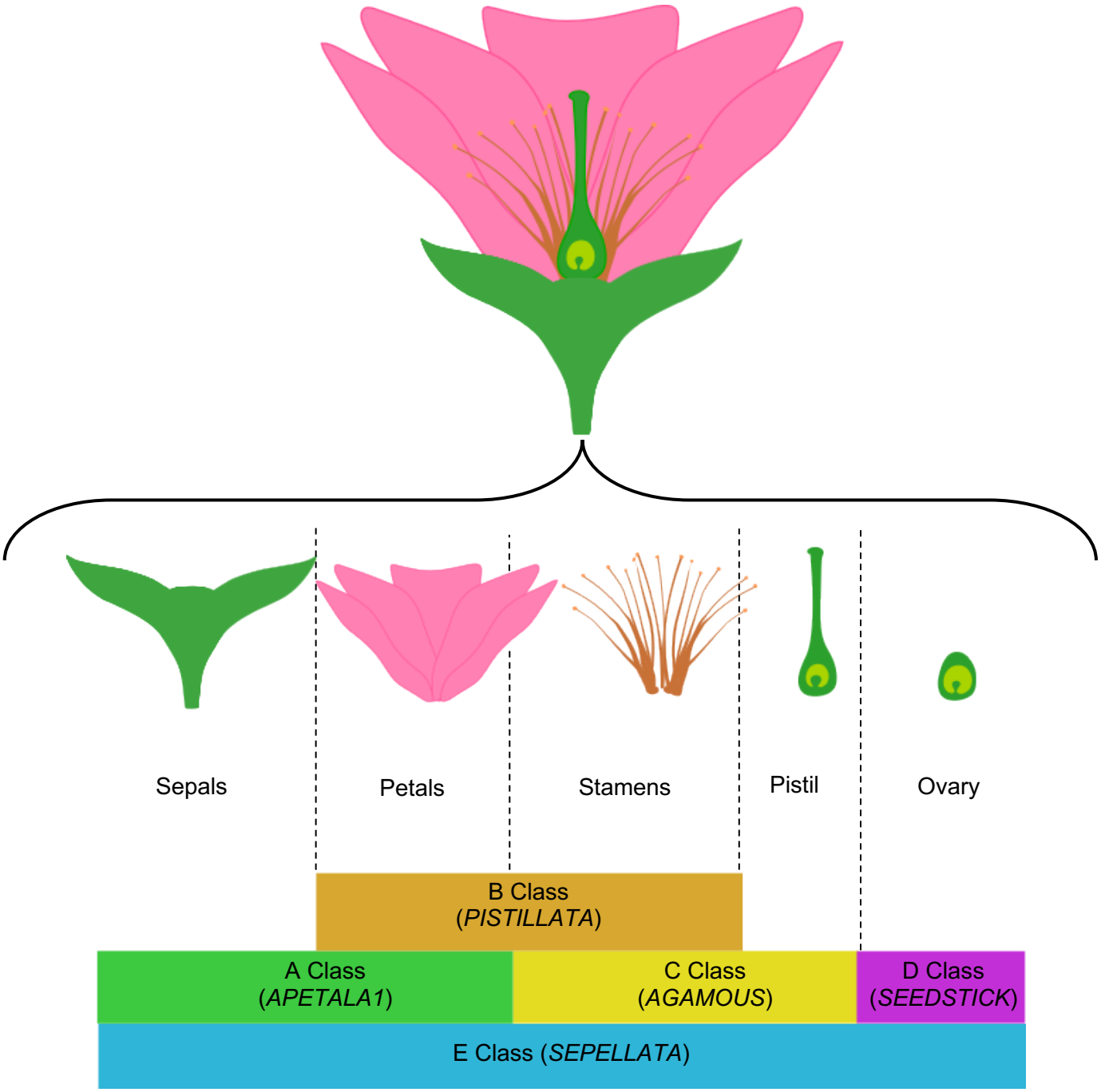


Figure 1.4: MADS-box gene categories responsible for different whorl development in the ABCDE floral model. The MADS-box genes encode for proteins that then interact forming tetramers that are responsible for the formation of different floral organs. Adapted from Chen *et al*., (2019).

The ‘ABCDE’ (formerly the ‘ABC’) model was characterised using the model organism *Arabidopsis* as well as *Anthirrhinum*, (Coen and Meyerowitz 1991; Soltis *et al*., 2007; Bowman and Moyroud 2024). This model suggests that the genes and the proteins they encode that are expressed specifically in one or a few whorls regulate and interact with one another to specify floral organ fate. An example of negative regulation in *Arabidopsis* is an A function MADS-box gene *APETALA1* (*AP1*) and a C function MADS-box gene *AGAMOUS* (*AG*). *AP1* is expressed throughout the early flowering bud but is negatively regulated by *AG* restricting it to the 1st and 2nd whorls whilst the *AG* protein functions in the 3rd and 4th whorls (Figure 1.4). This relationship in *Arabidopsis* showed that different whorls are specified partly due to the genes being expressed in discrete domains of the developing flower (Irish, 2017). However, some classes of MADS-box genes are absent from non-flowering plants. For example, the *SEPELLATA* class (Figure 1.4) that specify sepals are not found in gymnosperms, suggesting that MADS-box family expansion and gene duplication was critical for the evolution of floral morphology (Zahn *et al*., 2005). Interestingly, SEPELLATA genes do not show whorl-restricted expression (Figure 1.4). Further analysis of the functions of MADS-box genes present in non-flowering plants could help understand this family’s larger role in plant development.

1.6. The MADS-box gene superfamily

Transcription factors are proteins that regulate downstream gene expression by binding to DNA-regulatory sequences that may be upstream of a gene sequence or in introns (Yuan and Perry, 2011). The name ‘MADS’ is an acronym for the four genes in four diverse species used in the identification of the gene family, consisting of:

MINICHROMOSOME MAINTENANCE1 from *Saccharomyces*, AGAMOUS (AG) from *Arabidopsis thaliana*, DEFICIENS from *Antirrhinum majus*, and SERUM RESPONSE FACTOR (SRF) from *Homo sapiens* (Gramzow and Theißen, 2013).

The MADS-box family is known to have functional roles in plant architecture and development (Marchant *et al*., 2022). For example, changes in plant morphology have been noted due to MIKCc genes being expressed outside of floral organs, such as petal identity being transferred to organs outside of the flower in the dove tree (*Davidia involucrata*) (Vekemans *et al*., 2011). However, these functions lead to broad expression throughout different tissues and developmental stages.

The structure of MADS-box proteins has been used to categorise genes within the superfamily into specific types. Differences have arisen within the superfamily because of genome-wide duplication events happening multiple times, allowing the sequences of duplicates to diverge over time and enable further sub-categorisation (Litt and Irish, 2003). All MADS-box proteins contain a conserved DNA binding region (the MADS-box) after which they can be separated into two categories: Type-I and Type-II (Figure 1.5). The differences between these families can be best seen through differences in protein structure (Riechmann and Meyerowitz, 1997). In plants, Type-II MADS-box proteins contain a highly conserved N terminal MADS-box (M) DNA binding domain. The M-domain is approximately 55-60 residues long and has a sequence identity of 40% between *H. sapiens* and *Arabidopsis* (Lai *et al*., 2019). A moderately conserved Intervening (I) is involved in DNA-binding specificity (Zhang *et al*., 2017), although it does show less conservation (Kaufmann, Melzer and Theißen, 2005). A Keratin-like (K) domain is the second most conserved domain in a MADS-box protein and is

responsible for protein-protein interaction and dimerization (Gramzow and Theißen, 2013; Lai *et al*., 2019). A highly variable C-terminal (C) domain with few conserved secondary structural motifs completes the protein, and is involved in transcriptional activation (Lai *et al*., 2019).

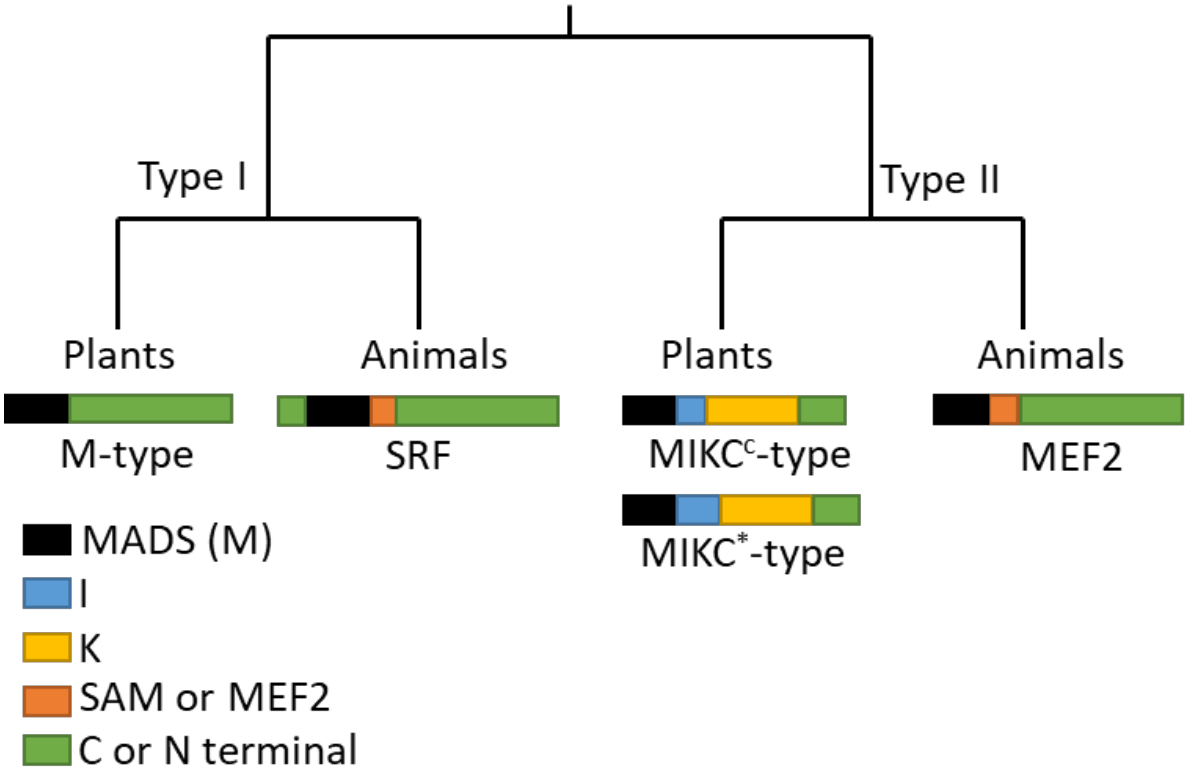


Figure 1.5: The domain structures present in Type-I and Type-II MADS-box genes in both plants and animals. The green represents the C-terminus domain in all but animal SRF proteins where it represents the C and N terminal. Fungi such as the yeast MADS-box proteins MCM1 hold similar structures to mammalian SRF and MEF2 proteins. Adapted from Nam *et al*., (2004).

In contrast, Type-I MADS-box proteins in plants do not contain a K-domain nor an I-domain, leading to a shorter total protein (Yang, Fanning and Jack, 2003) (Figure 1.5). The Type-II family is further categorised into MIKC^{c} (the classic model) and MIKC* (the deviant) proteins. Both these families contain the four domains described above

(M, I, K, C), (Kaufmann, Melzer and Theißen, 2005). However, MIKC^c and MIKC* are categorised through their differing structure, namely the differences in the I and K domains with MIKC^c I and K domains being shorter than MIKC* I and K domains, (Henschel *et al*., 2002) (Figure 1.5). Analysis of the number of introns present in these proteins shows that the MIKC^c family varies between 2 to 11 introns (Figure 1.6) whilst the MIKC* family varies between 8 to 16 introns. This pattern has been observed in *Pyrus*, *Arabidopsis*, and *Oryza sativa* (Meng *et al*., 2019). In comparison, Type-I genes possess no introns. Although they are all transcription factors, the different structures of the Type-I and Type-II MADS-box proteins leads them to have different specific functions.

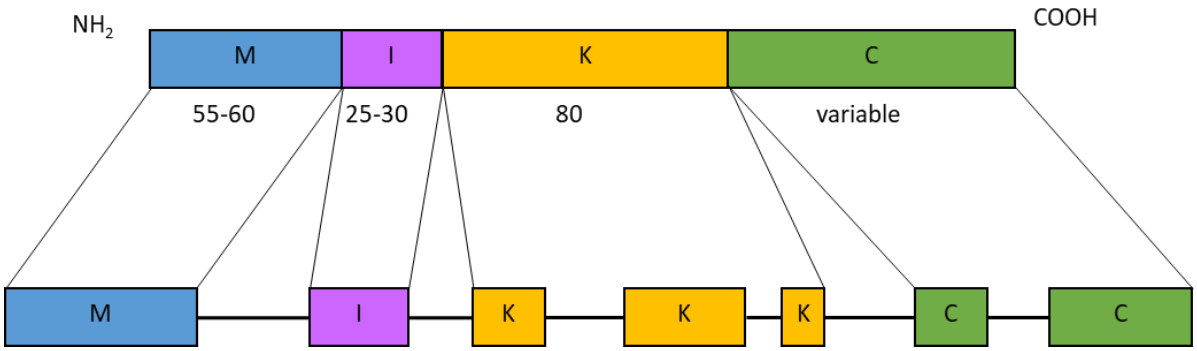


Figure 1.6: The MADS-box Type-II MIKC domain structure. The four domains MADS (M) DNA binding domain, Intervening (I) domain, Keratin-like (K) domain, and C-terminal (C) domain, the numbers correspond to amino acid length. The intron-exon structure of the MIKC^c protein CsMADS1 a cucumber homologue to *Arabidopsis* APETALA3 is shown, with introns being represented by lines. Adapted from Yang, Fanning and Jack, (2003), Tanabe *et al*., (2005), and Lai *et al*., (2019).

1.7. The evolutionary history of the MADS-box genes

The origin of MADS-box genes has been linked with deep evolutionary time. The M-domain was present in the early evolution of extant Eukaryotes. Basal protists and excavates do not have a full complement of MADS-box gene domains, perhaps due to a loss of MADS-box domains (Gramzow, Ritz and Theißen, 2010). It is thus considered likely that the common ancestor of all Eukaryotes contains a gene with an M-domain as MADS-box genes are present in plants, animals, and fungi (Gramzow, Ritz, and Theißen, 2010). It is believed that the last common ancestor of all Eukaryotes contained a single MADS-box gene, with a duplication event occurring before the divergence of plants and animals to produce two lineages of MADS-box genes, Type-I and Type-II (Alvarez-Buylla *et al*., 2000). Consistent with this, Type-I and Type-II MADS-box proteins can be found in plants, animals, and fungi (Figure 1.5). This is shown through a group of *Arabidopsis* Type-I MADS-box domains being closely related to the SRF- like sequences in animals (Messenguy and Dubois, 2003) (Figure 1.5). In chlorophytes and charophytes MADS-box genes cannot be separated into Type-I and Type-II categories as the genes present only contain an M-domain (Gramzow and Theissen, 2010; He *et al*., 2019). For example, the green unicellular alga *Chlamydomonas reinhardtii* and the highly reduced unicellular red alga *Cyanidioschyzon merolae* although they lack all domains, only showing the presence of an M-domain (Thangavel and Nayar, 2018). Despite this, it can be concluded that the most recent common ancestor (MRCA) of chlorophytes and streptophytes (which includes the land plants) contained a protein with a MADS-domain, and that the diversification of the MIKC^{C} and MIKC* Type-II MADS-box genes occurred after land

plants diverged from the *Chlamydomonas* lineage, (Thangavel and Nayar, 2018) (Figure 1.7).

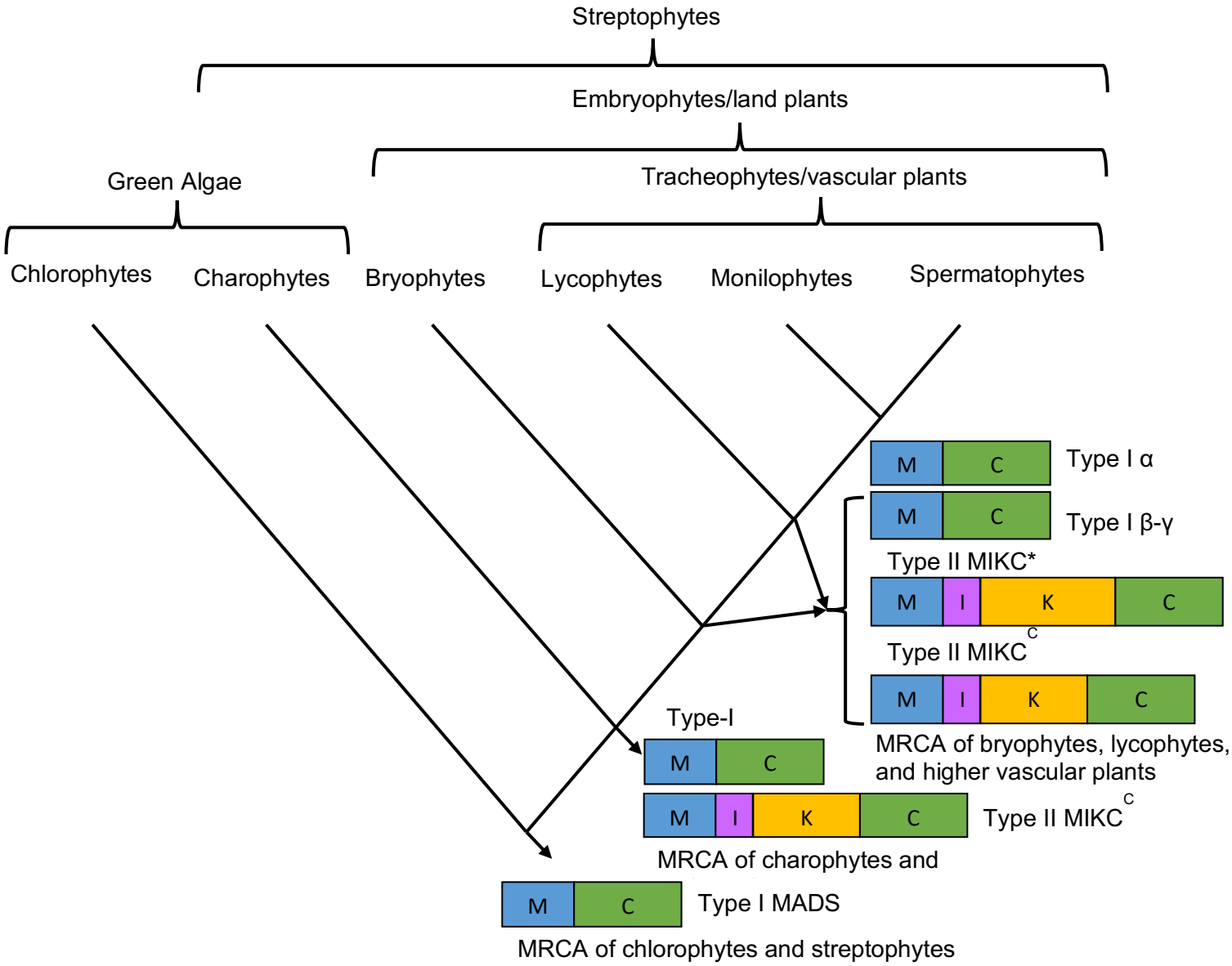


Figure 1.7: Evolutionary tree with the diversification events of the MADS-box family. The blue box represents the M-domain, purple the I-domain, yellow the K-domain, and green the C-terminus domain. Adapted from Thangavel and Nayar, (2018).

In green algae, the closest relatives to land plants, MADS-box gene function has been hypothesised to play a role in the haploid reproductive cell development (Thangavel and Nayar, 2018). This would correspond to the function of Type-I MADS-box genes in the gametophyte life cycle stage of angiosperms (Gramzow *et al*., 2012). The representative bryophyte *Physcomitrium* has Type-II MIKC^{c} MADS-box genes that have been shown to function in vegetative growth and reproduction (Koshimizu *et al*., 2018). Some genes have been co-opted into both moss gametophytes and angiosperm sporophytes in similar ways for example in root hair development (Jones and Dolan, 2012). Functional studies on lycophytes are limited. The model lycophyte *Selaginella* shows the conserved expression of MIKC* genes in gametophytes suggesting the MIKC* functions in gametophytes were conserved from early vascular plants (Kwantes *et al*., 2011). In gymnosperms the MADS-box gene family Type-II genes hold functions in the newly developed reproductive system of fruit development (Ng and Yanofsky, 2001). Throughout the evolution of land plants and their closest relatives what can be seen is conservation of MADS-box genes.

MADS-box gene families present in plants have expanded and diversified through whole genome duplication events as well as gene-specific evolution (Wang *et al*., 2024). There are vastly more MADS-box genes present in flowering plants compared with those present in animals and fungi (Theissen, Kim, and Saedler, 1996). Duplication events shaped the MADS-box family in plants into the known sub-categories, with gene duplication events and subsequent divergences being observed to alter the function and structure of genes for example the redundancy of *SEPALLATA* genes in *Arabidopsis* (Malcomber and Kellogg, 2005).

Type-I and Type-II genes show different birth-and-death rates with Type I exhibiting a higher birth-and-death rate, likely due to Type-I genes undergoing weaker purifying selection. Therefore, the death of a Type-I gene is likely to be less harmful than that of a Type-II gene, (Nam *et al*., 2004). This higher birth-and-death rate of Type-I genes alongside the fact Type-II genes are preferentially retained after a duplication event has meant Type-I clades are often young and fairly narrow-reaching whilst Type-II genes offer insight into ancient, conserved, and taxonomically widely distributed clades (Theißen and Gramzow, 2016). The shorter length and lack of introns of Type-I genes also allows them to be duplicated as functional units with more ease than Type-II genes (Gramzow and Theißen, 2013). Type-I MADS-box genes have been linked to female gametophyte development as well as the development of embryos and seeds in angiosperms, (Gramzow *et al*., 2012). Type-II MADS-box genes are further sub-categorised into MIKC^c and MIKC* families (Figure 1.7). Interestingly, some Type-II MIKC^c genes present in *Arabidopsis*, such as *AGL12* and *AGL25*-like genes do not contain a K-domain, suggesting this was acquired in the evolution of Type-II genes after these genes became integral to function (Henschel *et al*., 2002).

The observed increase in the diversity of MADS-box genes in plants is closely linked to the evolution of flowers, seeds, and fruits (Gramzow and Theißen, 2013). For example, the *APETALA1/FRUITFULL* family of MIKC^c MADS-box genes are only found in angiosperms, whilst the origin of genes responsible for male and female gametophytes predates the flower as these can be found in gymnosperms(Endress, 2001), for example putative AG genes in Gingko have been observed in ovule development (D'Apice *et al*., 2022). In addition, Type-II genes show duplication events and subsequent conservation of the genes that show specific functions relating to

novel seed and reproductive organs only present in spermatophytes (Kaufmann, Melzer and Theißen, 2005). In particular, it is known the MADS-box family is important in floral organogenesis, following the 'ABCDE' model to specify floral organ identity (see Section 1.5.). It is believed that each category of MADS-box gene in the 'ABCDE' model are products of duplication events, (Litt and Irish, 2003). For example, the *APETALA3* and *PISTILLATA,* two B-function genes present in *Arabidopsis* that are involved in stamen development, originate from a duplication event arising from a single ancestor gene lineage (Krizek and Meyerowitz, 1996). Not only have duplication and loss of gene events been linked to the evolution of more complex structures and diversification of flowers, but changes in the gene expression of floral homeotic MADS-box genes also show differences in the flower when altered (Mizukami and Ma, 1995). Therefore, when looking into the evolutionary development of plants, knowledge of the MADS-box genes is vital to understanding the origin of floral reproductive structures (Chen *et al*., 2019).

1.8. MADS-box protein structure and binding mechanism.

The different modular domains of MADS-box proteins have been found to have different functions. The highly conserved M-domain is present in both the Type-I and Type-II families, and is 55-60 amino acids in length (Yang, Fanning and Jack, 2003). The M-domain was identified as a highly conserved structure of 60 amino acids, and a crystal structure of the SRF core revealed the MADS-box structural motifs had high conservation between species as the orientation and fit of the coiled-coil MADS-box sequence remained constant along DNA recognition sites (Pellegrini, Tan and Richmond, 1995). At the amino end of the M-domain, DNA-binding occurs, whilst the

carboxyl end of the domain provides a surface – two anti-parallel β sheets - for dimerization (Tan and Richmond, 1998). Following the M-domain is the I-domain, previously named the linker (L) domain, for its placement between the two highly conserved domains of M and K. Few biochemical functions have been associated with this domain, although it has been suggested that it could be involved in the functional specificity of the MADS proteins, which has been established using domain swapping experiments, (Krizek and Meyerowitz, 1996; Riechmann and Meyerowitz, 1997a).

In Type-II proteins the third domain is the K domain, which is conserved at the amino acid secondary structure level and is 80 amino acids in length. The K domain forms a series of amphipathic α-helices (Riechmann and Meyerowitz, 1997b) and is important in protein-protein binding, with three sub-domains K1, K2, and K3, that form different sections of the α-helices and interhelical regions the 3 K introns form the K1, K2 and K3 domains (Ma, Yanofsky and Meyerowitz, 1991) (Figure 1.6). The conservation of the MIKC proteins can be seen through the similar intron patterns, with introns in between the domains as well as within the K domain bracketing the K1 and K3 subdomains, (Kaufmann, Melzer and Theißen, 2005) (figure 1.6). The fourth domain is the C-terminal domain at the carboxyl end of the protein both Type-I and Type-II. This domain is variable in length and not well-conserved (Theißen, Kim and Saedler, 1996). However, some MADS proteins have a transcriptional activation domain within the C-domain whilst other studies suggest the C-domain acts as a mediator in the formation of ternary complexes with other MADS-box proteins to form activation domains (Parenicova *et al*., 2003). Ternary complexes are formed when three different molecules bind together and they can alter the specificity and function of a protein (Egea-Cortines *et al*., 1999).

MADS-box proteins are generally combinatorial transcription factors that require binding between two or more proteins to function and can form multi-protein DNA-bound complexes through associating in a specific way, (Wolberger, 1998). As such, homodimerization (binding to another copy of the same protein) and heterodimerization (binding to a different protein) of the proteins will determine the specific regulation of the gene targets (Messenguy and Dubois, 2003). This protein-protein binding has increased in complexity within the 'ABCDE' model, leading to the formation of 'floral quartets' comprising tetrameric MADS domain protein complexes, (Puranik *et al*., 2014) (Figure 1.8). The M-domain and the I-domain are involved in specifying certain protein-protein interactions and are used when interacting with a co-factor, for example when forming multimeric complexes with E-class proteins (Krizek and Meyerowitz, 1996). The presence of a K-domain in Type-II proteins provides interaction surfaces for protein-protein binding, which allows for homodimerization, and heterodimerisation, as well as the tetradimerisation seen in the 'floral quartet' (Mason and Arndt, 2004). The K domain has a coiled-coil structure that acts as a leucine zipper this allows for interactions between different proteins, (Riechmann and Meyerowitz, 1997b) and can alter the interaction patterns of Type-II proteins, increasing the potential regulatory and functional diversity of these transcription factors (Lai *et al*., 2019). Each domain within a MADS protein thus has a different function and the individual proteins must work co-operatively to have an effect, as such it is important for them to be able to bind efficiently and in the correct combinations.

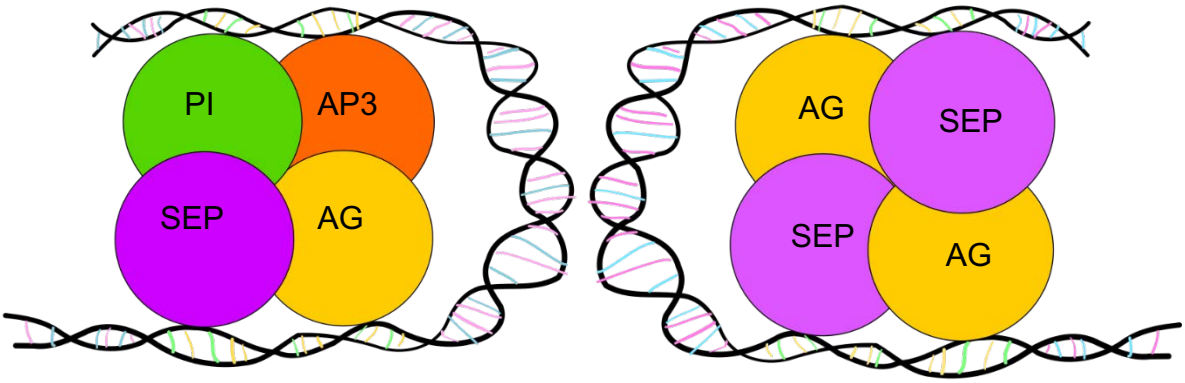


Figure 1.8: Examples of MADS-box tetramers on DNA, known as 'floral quartets'. The left floral quartet contains two B-class (PI and AP3), one C-class (AG), and one E-class (SEP) genes encoding mRNA for transcription factors regulating stamen development. The right quartet contain two C-class (AG) and two E-class (SEP) genes encoding carpels. Figure 1.4 for ABCDE model. Adapted from Ruelens *et al*., (2017).

MADS-box proteins accomplish their tasks by forming higher order complexes, such as the tetramers of floral quartets, which then act on promoter sequences of their target genes. MADS-box proteins of plants are known to bind DNA only after dimerization, (Ruokolainen *et al*., 2010). MADS-box proteins bind to a sequence of DNA which has the consensus sequence $CC(A/T)_6GG$ known as a CArG-box, this DNA-binding domain acts to regulate the transcription factors of the MADS-box family (Aerts *et al*., 2018). This regulation is achieved through different affinities of DNA-binding dependent upon of the protein-protein complex the MADS-box proteins are in, with higher affinity noted to protein-protein dimers (Messenguy and Dubois, 2003). MADS-domain proteins are much more likely to bind in a heterodimer/hetero-tetramer, a hydrophobic kink region within the K domain prevents self-association during oligodimerisation, increasing the diversity of complexes formed (Puranik *et al*., 2014). This is observed when highly similar DNA binding motifs such as a CArG-box are

present in multiple datasets, suggesting there is a masking of more specific sequences by the vast combination of heterodimers (Aerts *et al*., 2018).

Structural studies of MADS-box proteins are largely limited to animal species, with crystal structures being resolved for a human SRF and MEF type MADS-box protein, a mouse MEF type protein, and the budding yeast MADS-box protein MCM1 (Qiu *et al*., 2023). Only one plant MADS-box protein has had its crystal structure resolved the *Arabidopsis* MADS-box protein SEP3 (Lai *et al*., 2019). Although predicted protein structures have been generated, there have been differences identified between predictions and resolved structures (Qiu *et al*., 2023) (Chapter 5). MADS-box interactions have also been characterised in the small MADS-box family found in the moss *Physcomitrium* (formerly *Physcomitrella*) (Henschel *et al*., 2002). However, in between the early land plant example and the angiosperm examples, functions of the MADS-box genes have yet to be determined. This is a vital gap in knowledge in understanding the evolution of MADS-box genes in relation to seed and flower evolution.

1.9. Using *Ceratopteris* to investigate the evolution of the MADS-box family

As ferns lie in the sister clade to seed plants (Figure 1.1), they present an opportunity to investigate land plant development at this crucial stage in evolution, when large plants present on dry land and were changing their reproductive systems to adapt accordingly after already having co-opted a vascular system (Bui *et al*., 2015). Ferns generally have long life spans, therefore finding a model organism within this family

has been challenging. The massive genomes associated with ferns, due to multiple duplication events throughout their evolution, are an issue when it comes to carrying out molecular-genetic research. Heterospory is important in seed formation as there is a reliance on small pollen grains and a large ovule. Some heterosporous ferns have relatively small (<1GB) genomes that have allowed for sequencing, with the heterosporous ferns *Azolla filiculoides* (750MB) and *Salvinia cucullate* (260MB) now having their genome sequenced in their entirety (Banks, 2018; Li *et al* 2018) with nine putative MADS-box genes being identified in *Azolla filiculoides* (Qui and Yu, 2003).

Not only does the size of fern genomes present difficulties, but the large size of the genes within the genomes is yet another hurdle to overcome with traditional cloning and transformation methods having a low efficiency (Plackett *et al*., 2014). *Ceratopteris richardii* (referred to as *Ceratopteris* throughout this work) has had its genome (11.25 Gbp, n= 39) sequenced and annotated, providing the first homosporous vascular plant reference genome (Marchant *et al*., 2019). The publication of the *Ceratopteris* genome (Marchant *et al* 2019) and its update in 2022 (Marchant *et al*., 2022) overcame some barriers to molecular-genetic research by providing a full set of annotated protein-coding genes, promoter regions, etc. Since then, another homosporous fern genome, that of *Adiantum capillus-veneris* has also been sequenced (Fang *et al* 2022) along with the genome of a tree fern (Huang *et al* 2022).

Ceratopteris is a diploid fern, that can be transformed with relative ease when compared to other ferns (Stuessy and Weiss-Schneeweiss, 2019) (Figure 1.9). Unlike many ferns, Ceratopteris has a short life cycle and can be grown under laboratory

conditions (Kinosian and Wolf, 2022). Moreover, *Ceratopteris* has a free-living haploid gametophyte life-cycle stage that can be used for transformation (Plackett *et al*., 2014).

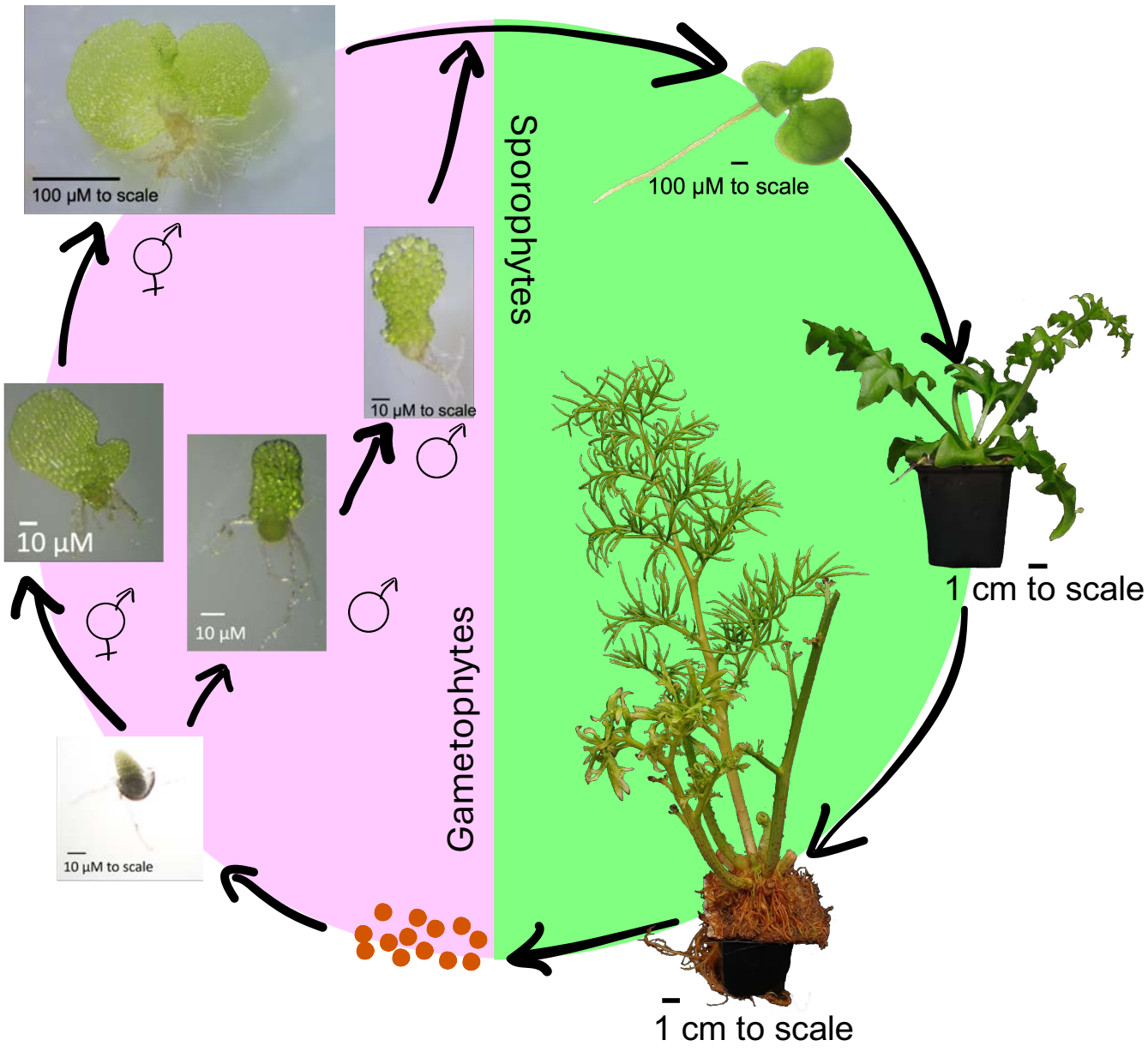


Figure 1.9: *Ceratopteris* life cycle. The free-living gametophyte stage is highlighted in pink with a life span of approximately 4 weeks, the sporophyte stage is highlighted in green with a life span of approximately 20 weeks. Figure 1.2 gives a comparison to the flowering plant life cycle and the greatly reduced gametophyte stage

Monilophytes are the closest relatives to seed bearing plants (Knie *et al*., 2015). This places *Ceratopteris* at a vital point in evolution when investigating the evolutionary changes in reproductive systems (Linkies *et al*., 2010). *Ceratopteris* has been used as a model organism for over three decades, and the recent advent of the high-quality genome will enable such investigations (Kinosian and Wolf, 2022).The MADS-box gene family in *Ceratopteris* was annotated in the published genome (Marchant *et al*., 2022). This along with the previously established molecular methods associated with *Ceratopteris* make it an ideal organism to use when investigating the MADS-box genes and their role in the evolution of seed and flower base reproductive systems.

1.10 Aims and objectives

1.10.1. Aim

The overarching aim of this thesis was to determine the function of MADS-box genes in Ceratopteris, within the sister clade to spermatophytes. To achieve this aim, the following specific objectives will be addressed in the following chapters.

1.10.2. Objectives

(i) Identification and validation of the full complement of MADS-box genes present in the fern *Ceratopteris* using bioinformatic, phylogenetic and molecular cloning approaches (Chapter 3).

(ii) Analysis of potential functions for *Ceratopteris* MADS-box genes via expression analysis using RT-PCR and qPCR on tissue from different stages of development (Chapter 3.2).

(iii) Generation of transgenic RNAi knockdown lines of a selected *Ceratopteris* MADS-box protein to define for the first time the function of a non-seed vascular plant MADS-box gene, via analysing loss/reduction of function (Chapter 4.2).

(iv) Generation of overexpression lines of a selected *Ceratopteris* MADS-box gene in both *Ceratopteris* and *Arabidopsis* to determine and compare gene function(s) (Chapter 4.3).

(v) Prediction of *Ceratopteris* MADS-box protein function and interactions via comparative modelling of fern and *Arabidopsis* MADS box proteins (Chapter 5.2, 5.3, and 5.4).

(vi) Investigation of protein-protein interactions within the fern MADS-box family using Yeast-II-Hybrid assays of *Ceratopteris* MADS-box proteins (Chapter 5.5).

CHAPTER 2: MATERIALS AND METHODS

2.1. Plant materials

Wild type (WT) spores of *Ceratopteris richardii* Hn-n strain (https://c-fern.org/) were obtained from Dr Andrew Plackett (University of Birmingham). Wild type (WT) seeds of *Arabidopsis thaliana* Col-0 were used, which can be obtained from the European *Arabidopsis* Stock Centre (https://arabidopsis.info/).

2.1.1. Preparation of growth media

1X C. fern medium (Appendix) was prepared as in Plackett *et al*. (2015) adapted from Klekowski (1969). For plates approximately 25 mL of 1X C. fern media with 1% agar, adjusted to pH 6.0, was used per 90 mm petri dish, poured in a sterile laminar flow hood. Plates were to be used on the day and not stored to avoid drying out or contamination. Sterile dH_2O was added consistently every other day to keep plates moist, in a sterile flow hood.

For antibiotic selection and hormone treated plates, the medium was first cooled to 55 °C before the required antibiotic or hormone was added and the plates poured as above. For ease of transplant, harvest, and observation of gametophytes, 90 mm petri dishes with set C. fern medium were overlayed with cellophane discs, that had been sterilised between filter paper discs (Whatman plc) in a glass petri dish for 20 minutes at 121 °C, using sterile Watchmaker forceps in a laminar flow hood.

For sterile sporophyte growth Magenta pots (Sigma-Aldrich) 77 mm x 77 mm x 97 mm were first autoclaved and then used, half-filled with C. fern medium, added in sterile laminar flow hood.

Plants grown on soil (Levington Peat Free All-Purpose compost), had 500 mL 1X C. fern media without agar added, to water the tray at the initial transplantation stage, and again if plants showed nutrient stress through yellowing as the grew.

2.1.2. Tissue Culture and growth conditions of *Ceratopteris richardii*

Ceratopteris spores were imbibed, cleaned, and stratified. *Ceratopteris* spores were incubated in 20% bleach (800 µL dH_2O, 200 µL NaClO, and 5 µL Tween20) at room temperature for 15 minutes whilst mixing gently. The spores were then washed x4 with sterile dH_2O before being resuspended in 1 mL dH_2O and incubated at room temperature in darkness for 48-72 hours. Spores were then plates on C. fern media plates.

Gametophyte plates of *Ceratopteris* were incubated at 28 °C with 50% humidity. Once on soil, sporophytes were incubated at 28 °C with 80% humidity. Both incubations had a long day/night cycle with 16 hours light and 8 hours dark cycles (Plackett *et al*., 2015).

2.1.3. Growth of Ceratopteris for RNA extraction

Wild type Hn-n *Ceratopteris richardii* was grown following the protocol described in Plackett *et al*., (2015). Tissue was harvested at immature and mature gametophyte

stages, as well as vegetative and reproductive sporophyte tissue at the following three stages: fiddle head, expanding fronds, mature fronds. This was to generate tissue for RNA extraction.

2.1.4. Generation of callus for transgenic fern generation:

Once sporophytes had developed a visible shoot and root (approximately two weeks of growth after fertilization) callus was induced by transplanting sporophytes to 90 mm petri dishes with 1X C. fern media with 1% agar and 5 μM 6-Benzylaminopurine (BAP) using 2 pairs of watchmaker forceps to remove any excess media and anchor roots in new media. Approximately 1 cm of space was left between sporophytes to allow room for growth. Sterile dH_2O was added every 2-3 days to keep plates moist.

Once callus tissue had developed (after approximately one to two weeks) it was transplanted to 90 mm petri dishes with 1X Murashige and Skoog (MS) medium containing 2% sucrose and 0.7% agar with 5 μM BAP media (MS-BAP), adjusted to a pH of 5.8. Using two pairs of watchmaker forceps, callus tissue was gently pressed into the media leaving around 1 cm between individual callus. Plates were sealed with Parafilm™ to maintain a moist environment. Once callus had grown (after approximately one week), each piece was broken into smaller pieces using watchmaker forceps and transplanted to new MS-BAP plates. After 14 days, callus was ready for DNA microparticle bombardment and was transplanted to 50 mm petri dishes containing 1X MS 2% sucrose 0.7% agar (pH 5.8) with 5 μM Kinetin.

2.1.5. Microparticle preparation for biolistic bombardment:

The callus was then used for transformation using a biolistic gun (PDS-1000/He™ System, Bio-Rad). This was performed by sterilising and coating approximately 20 mg of gold microparticles (1.0 µm gold microcarriers – Bio-Rad).

2.1.6. Microparticles sterilisation

Microparticles were sterilised by incubating in 1 ml of 70% ethanol for 15 minutes at room temperature. The ethanol was then removed by pulse centrifuging and 1 ml of sterile dH_2O was added. Microparticles were then shaken for 1 minute at approximately 200 rpm, pulse centrifuged to collect the microparticles in a pellet and 1 ml of sterile dH_2O was added. Microparticles were washed a total of three times. Microparticles were then suspended in 50% glycerol to a concentration of 60 mg microparticles per ml assuming no loss during sterilisation.

2.1.7. Microparticle coating

Microparticles were coated with the required DNA construct. Two aliquots of microparticles were needed for each round of bombardment to allow for a no-DNA control. Whilst constantly shaking, 50 µL microparticles were aliquoted into each 1.5 mL Eppendorf tube. In quick succession 5 µL DNA construct, 50 µL $CaCl_2$, and 20 µL spermidine were added and the contents shaken for a further 5 minutes on a vortex mixer with a tube holder attachment. For the no-DNA control 5 µL of sterile dH_2O was

added in place of the DNA. Microparticles were then washed twice with 140 µL 70% ethanol and suspended in 50 µL 100% ethanol.

2.1.8. Biolistic bombardment of *Ceratopteris* callus:

Once microparticles were ready for bombardment the biolistic gun was prepared. In an ESCO Laminar Flow Cabinet equipment was sterilised in 100% ethanol for 20 minutes. This included: 2x bayonet forceps, 1x Watchmaker forceps, the loading stage, cap, and firing cylinder from the PDS-1000/He^{TM} System, 4x Bio-Rad Biolistic Macrocarrier holders, 4x metal cartridges, 4x circular wire mesh discs. 4 x P/N 9202964 “C” Biolistic rupture disks 1,100 psi were prepared in isopropanol, these were allowed to dry before use. The PDS-1000/He^{TM} System gun was cleaned with 70% ethanol when placed within the laminar flow cabinet. Bayonet forceps were used to insert microcarrier discs into each cartridge, concave side down, and the discs were pressed down until discs they lay flat within the cartridge. 10 µL of coated microparticles were pipetted onto the centre of each macrocarrier and allowed to dry.

Watchmaker forceps were used to place the rupture discs in the firing cylinder, ensuring they lay flat. Using a torque wrench, the firing cylinder was placed into the gun and tightened. Using Bayonet forceps, a wire mesh was placed over the hole in the firing stage and the loaded bombardment cartridge was then placed on top, disc side down. The loading stage was then placed above the sample stage in the gun and the callus plate placed on the sample stage ensuring the plate was centred and the lid had been removed. The He vacuum was filled to 27.5 psi in He vacuum and held until the gun reached approximately 1000 psi. This process was repeated with a new

macrocarrier each time. The wire mesh disc could be used for a total of 4 bombardments and would need replacing if more were being performed. The no-DNA control was bombarded first with three more bombardments being performed for the DNA construct. Three separate populations were then bombarded in this manner, giving 9 total DNA-transformed callus populations and 3 no-DNA controls. Regenerated plants were grown on soil and spores were collected. Germinating spores then underwent three rounds of antibiotic selection on MS-plates before being transplanted to soil where 4 plants from each resistant line were grown for phenotyping and genotyping and T2 spore collection.

2.2. *Ceratopteris* genotyping and phenotyping of RNAi CMADS1 knockdown T1 lines

2.2.1. Genotyping

cDNA of each transformant line plus the empty vector control lines and the wild-type control lines was used in a qPCR analysis to recognise any knockdown of CMADS1 present (see chapter 4.2.1).

2.2.2. Phenotyping

One frond was harvested from each plant of each T1 *Ceratopteris* line (Figure 2.1). This frond was photographed and measured using a ruler by eye. Branches per frond and pinnae per frond were then counted by eye. Three pinnae from each frond were cut and used for further phenotyping (Figure 2.1). A dissecting microscope (G X

Microscopes by GT Vision Ltd ultraZOOM-2) was used under x0.8 magnification to view pinnae and sporangia of the transgenic plants. To view sporangia pinnae were dissected using x2 watchmaker forceps to allow for photography and counting of enclosed sporangia. Pinnae that were consistently re-curling were fixed in 100% ethanol.

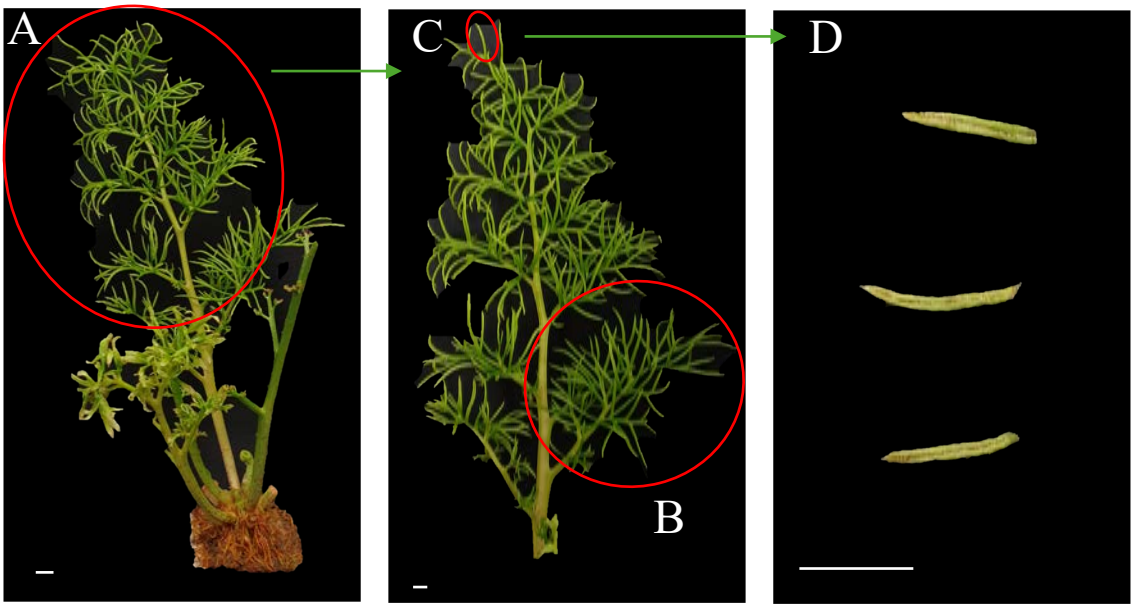


Figure 2.1: Measurements taken for phenotyping of *Ceratopteris* knockdown lines.
The red circle labelled A shows an example frond taken for measurement; frond length measurement taken from base to tip. The red circle labelled B shows one branch, while the red circle C shows a pinna. Panel D is a close-up image of an example three pinnae chosen for measurements of sporangia. Scale bars represent 1 cm.

2.3. *Arabidopsis thaliana*

2.3.1. Preparation of growth media

½ strength MS medium with 2% sucrose and 0.7% agar medium and the pH adjusted to pH 5.8 (Appendix 1). Approximately 25 mL of ½ MS 2% sucrose 0.7% agar medium

was used per 90 mm petri dish. Plates were to be used on the day of preparation and not stored, to avoid contamination.

For antibiotic selection and hormone treatment plates, the ½ MS medium was first cooled to 55 °C before the required antibiotic or hormone was added and the plates poured. For T1 selection of transformed *A. thaliana* lines 50 µg/mL Kanamycin was included in the plates.

When growing on soil a composite of ratio 4:2:1 compost (Levington f2):vermiculite:perlite was used.

2.3.2. Growth conditions

For imbibition and surface sterilisation, *Arabidopsis* seeds were incubated for 5 minutes in 70% ethanol at room temperature, followed by a 15-minute incubation in 10% bleach (900 µL dH_2O, 100 µL NaClO, and 5 µL Tween20) at room temperature whilst mixing gently. The seeds were then washed x4 with sterile dH_2O before being sown to either soil or plates depending on the selection necessary. Before planting seeds on plates, imbibed and washed *Arabidopsis* seeds were stratified by incubating at 5 °C for 48 hours in darkness. *Arabidopsis* plates were incubated at 25 °C with a 16 hour light cycle.

At approximately 7-14 days old when four true leaves were present, *Arabidopsis* seedling roots were developed enough to be moved to soil without damage whilst not being too embedded in the media, which would lead to breakages. Thus, by 7-14 days

old the seedlings should survive the move from the nutrient rich and high humidity environment of the growth media to soil. Soil mixture was placed into pots ensuring no large air gaps. A tray was filled approximately halfway with water and the pots placed into it. When the soil was saturated with water, the tray was emptied, and the seedlings transplanted into moist soil. A plastic lid was used cover pots for a few days to ensure successful transplants. Pots were then placed into a growth cabinet with a 16-hour light cycle, at 25 °C.

After one week in soil, *Arabidopsis* seedlings had 1-2 leaves harvested, flash frozen, and stored at -80C for DNA extraction and genotyping. After 6-9 weeks on soil, depending on floral cluster production, two young floral clusters were harvested for immediate phenotyping and a further two young floral clusters were harvested and snap frozen for future RNA extraction.

2.3.3. *Arabidopsis* genotyping and phenotyping of overexpression *35S::AG* and *35S::CMADS1* T1 lines

2.3.4. Phenotyping

Young floral clusters were observed under a dissecting microscope at x2 magnification (G X Microscopes by GT Vision Ltd ultraZOOM-2) with sepals, petals, and stamens being counted before being dissected for photography.

2.4. *Ceratopteris* genome validation and *Ceratopteris* MADS-box cloning.

2.4.1. cDNA isolation:

Complementary DNA was generated from gametophyte tissue at 5, 7, 10, and 14 days of growth and from vegetative fiddle head, expanding fronds, and mature fronds, and reproductive fiddle head, expanding fronds, and mature frond sporophyte tissue stages. Tissue was harvested using 2x watchmaker forceps under a dissecting microscope and tissue was placed into a 1.5 ml Eppendorf tube before snap freezing using liquid nitrogen. RNA extraction was then performed using RNeasy Plant Mini Kit (Qiagen) (see table X). For use in qPCR RNA samples should ideally be above 50 ng/μL so synthesised cDNA has a sufficiently high concentration (Ma *et al*., 2021). As such, multiple extractions were needed to gain samples above the concentration threshold in some cases. RNA was then used for cDNA synthesis using First-Strand cDNA synthesis using SuperScript II RT (Thermo Fisher Scientific) according to manufacturer's instructions.

2.4.2. Phusion PCR amplification and Gene Validation:

Using annotated sequences from *Ceratopteris richardii* genome (Marchant *et al*., 2019) and Dr Andrew Plackett's preliminary RNA sequencing data (unpublished), primers were designed to amplify the full-length predicted coding regions of each gene. Each primer was designed to have a length of between 15-25 base pairs, and a GC content of between 40%-60%, avoiding sequences with long runs of the same

base (see tables 1 and 2). Primers were designed so annealing temperatures for both forwards and reverse were designed to be similar in temperature at around 55 °C. As PCR with Phusion DNA polymerase (Thermo Fisher) was used, 3 °C was added to the Tm giving an annealing temperature of Tm + 3 °C. An extension temperature was chosen depending on the length of the sequence being amplified, with Phusion polymerase requiring 30 seconds per 1 kb.

PCRs were performed using a LifeECO, TC-96/G/H(b)C, Hangzhou Bioer Technology Co. Ltd PCR machine. The initial denaturation stage was 98 °C for 30 seconds, followed by 29 cycles of a 10 second 98 °C denature stage, 20 seconds of the chosen annealing temperature (plus 3 °C) depending on the primers being used, a 72 °C extension step of the chosen time based on product length (30 seconds per 1 kb). A final extension stage was added of 72 °C for 6 minutes.

2.5. Quantitative expression analysis

2.5.1. Quantitative RT-PCR

qRT-PCR was performed on an AriaMx machine (Agilent) using Brilliant III ultra-fast SYBR® Green low ROX qPCR master mix (Cat. 600892, Agilent). 384 well plates were used with each 5 µL reaction containing 2.5 µL Sybr Green Supermix, 0.25 µL of the appropriate forward primer at 10 µM, 0.25 µL of the appropriate reverse primer at 10 µM, and 2 µL of the appropriate cDNA at 5 ng/µL. Four technical repeats were performed for each sample. The cycling parameters were: An initial denaturation stage of 95 °C for 10 minutes followed by 40 amplification cycles of a 15 second 95 °C

denaturation step, 20 second 60 °C annealing step, due to the short amplicon length elongation occurs during the cycle up to the denaturation temperature. A melt curve was then produced with a 15 second 95 °C step, 60 second 60 °C step, and a final 15 second 95 °C step with a ramp rate of 0.075 °C every second to produce a dissociation curve.

2.5.2. qRT-PCR data analysis

The normalised fold change in relative expression of each *CMADS1* knockdown line was calculated relative to *Ceratopteris* wild type, the $2^{-\Delta\Delta CT}$ method described in Livak and Schmittgen, (2001). CT values were normalised against three housekeeping genes, Actin (ActinB), a putative Ubiquitin gene, and a putative TATA gene (Ganger *et al*., 2015). Error bars represent upper and lower fold-change range and were calculated using standard deviation.

2.5.3. Agarose gel electrophoresis

0.8-1.0% agarose was used in 1xTAE (diluted from a 50x stock: 242 g Tris Base, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA pH 8.0, volume adjusted to 1 L using dH_2O). Depending on the number of samples being run either 50 mL or 100 mL TAE agarose solution was poured into gel trays with the appropriate gel combs. 6X gel loading dye (NEB) containing 1:1500 GelRed nucleic acid stain (Biotium) was added to the DNA samples before loading them into the set gel. A 1 kB DNA ladder (NEB) with 6X gel loading dye (NEB) added was used to help visualise the sample bands. All samples were loaded into the gel wells and run between 70V-90V at 120 mA for 45-

60 minutes using a Cleaver Scientific electrophoresis battery. DNA bands were visualised using a ChemiDoc MP imaging system on a Bio-Rad gel doc.

2.5.4. DNA extraction and purification

PCR products that were then run on agarose electrophoresis gels were extracted and purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research) following the manufacturers protocol. The band of the correct size, determined through the use of a 1 kb DNA gene ladder (NEB) was extracted using a UV light box and a razor blade. This was then placed into 1.5 mL Eppendorf tubes where the gel was dissolved in a high ionic strength buffer at 50C for approximately 10 minutes depending upon the weight of the band. Using the provided spin columns that bind DNA to a silica membrane the excised and dissolved DNA was washed using ethanol-based buffers and eluted into ≥ 6 µL dH_2O.

2.5.5. Ligation

Purified PCR product was ligated with the blunt ended pJET 2.1 cloning vector using the enzyme T4 DNA ligase, both provided in the CloneJet PCR Cloning Kit (Thermo Fisher Scientific). The solution incubated at room temperature for between 5-15 minutes for ligation to occur.

2.5.6. Plasmid purification from *E. coli* cells.

A QIAprep Spin Mini-prep kit (Qiagen) was used for plasmid purification according to the manufacturer's instructions. Plasmid DNA concentrations were determined using

a nanodrop spectrophotometer. The purified plasmid was sent for sequencing (Source Bioscience, Cambridge)

2.5.7. Restriction enzyme digestion

Restriction enzyme diagnostic digests were designed and performed depending on the sequence being cloned, using enzymes from New England Biolabs according to the manufacturer's instructions.

2.5.8. Sequencing

Sanger sequencing was performed using the Source Bioscience online rapid sequencing portal (https://genomics.sourcebioscience.com/). For each 5 µL sequencing reaction, a plasmid concentration of 100 ng/µL and a primer concentration of 5 mmol were used.

2.5.9. Expression analysis

RT-PCR as described in 2.3.2. was used to determine the absence/presence of MADS-box Type-II MIKC^c genes in different developmental stages.

2.6. Cloning of *CMADS1* and other genes in *Ceratopteris*, *Arabidopsis*, and *S. cerevisiae*

2.6.1. Vectors

pART7 (Gleave, 1992)

The 5 kB plasmid contains a CAMV 35S promoter and an Ampicillin resistance gene to allow for restriction cloning and donation of the overexpression promotor region.

pANDA (Miki and Shimamoto, 2004)

The 6.7 kB plasmid contains a maize ubiquitin promoter and an *E. coli gus* linker between the two inverted insertion repeats to allow for use in RNAi. The plasmid contains Kanamycin resistance in bacteria and Hygromycin resistance in plant.

pBOMBER (Plackett, Rabbinowitsch and Langdale, 2015)

The 11.5 kB plasmid contains a NOS promoter site and holds a spectinomycin resistance.

Gateway™ pDONR207™ Vector (ThermoFisher Scientific)

The 5.6 kB plasmid contains a gentamycin resistance cassette for selection in bacteria and AttP sites to allow for directional cloning of PCR products via BP recombination

pGADT7 (Clontech)

The 8 kB plasmid contains the GAL4 activation domain (AD) and ADH1 promoter region upstream of the MCS allowing for GAL4-AD protein fusions in yeast The T7 promoter and HA tag allow *in vitro* synthesis of N-terminal epitope tagged proteins. The plasmid contains Ampicillin resistance.

pGBKT7 (Clontech)

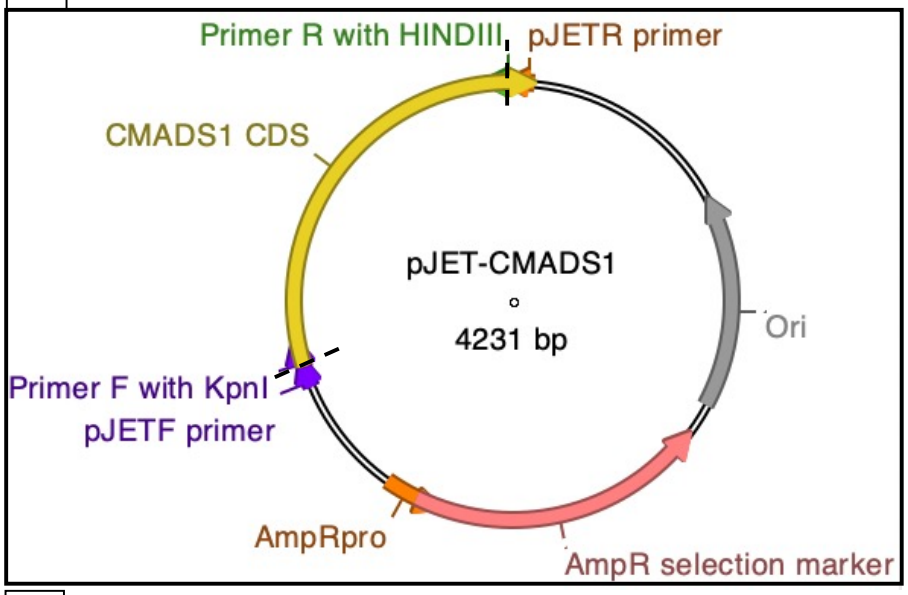
The 7.3 kB plasmid contains a GAL4 binding domain (BD) under the control of the ADH1 promoter upstream of the MCS this allows for GAL4-BD protein fusion in yeast.

The T7 promoter and HA tag allow *in vitro* synthesis of N-terminal epitope tagged proteins. The plasmid contains a kanamycin resistance.

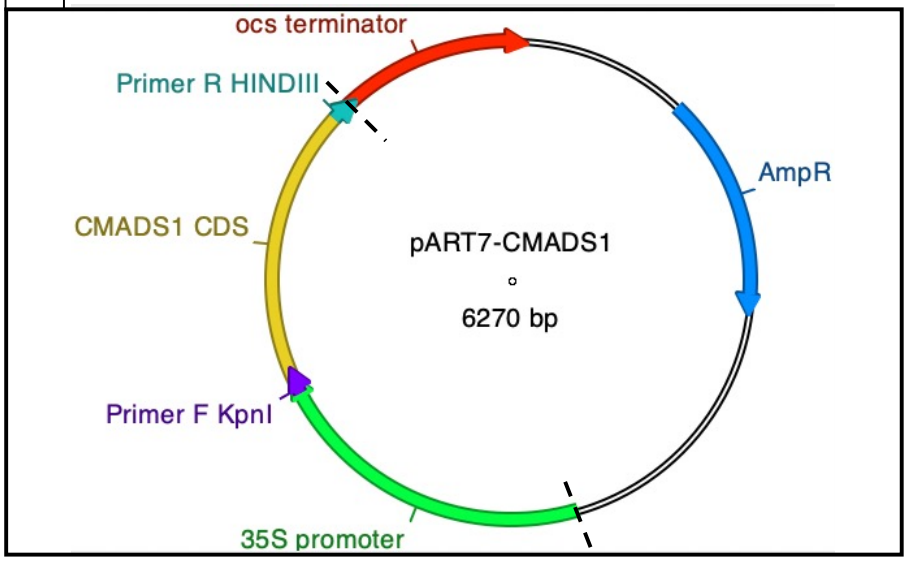
2.6.2. Vector construction.

For construct production restriction cloning was used with various restriction sites utilised.

A


Primer R with HINDIII
pJETR primer
CMADS1 CDS
pJET-CMADS1
4231 bp
Ori
Primer F with KpnI
pJETF primer
AmpRpro
AmpR selection marker

B


ocs terminator
Primer R HINDIII
AmpR
CMADS1 CDS
pART7-CMADS1
6270 bp
Primer F KpnI
35S promoter

C

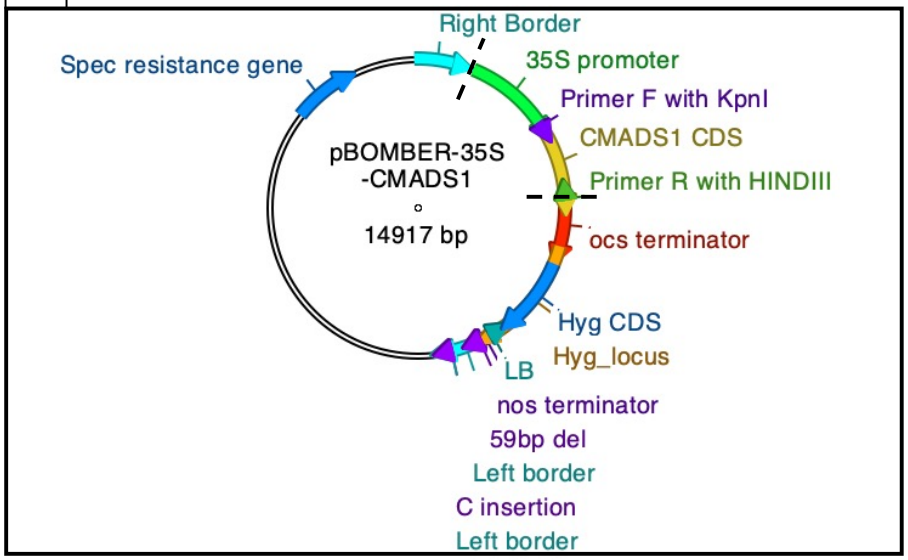
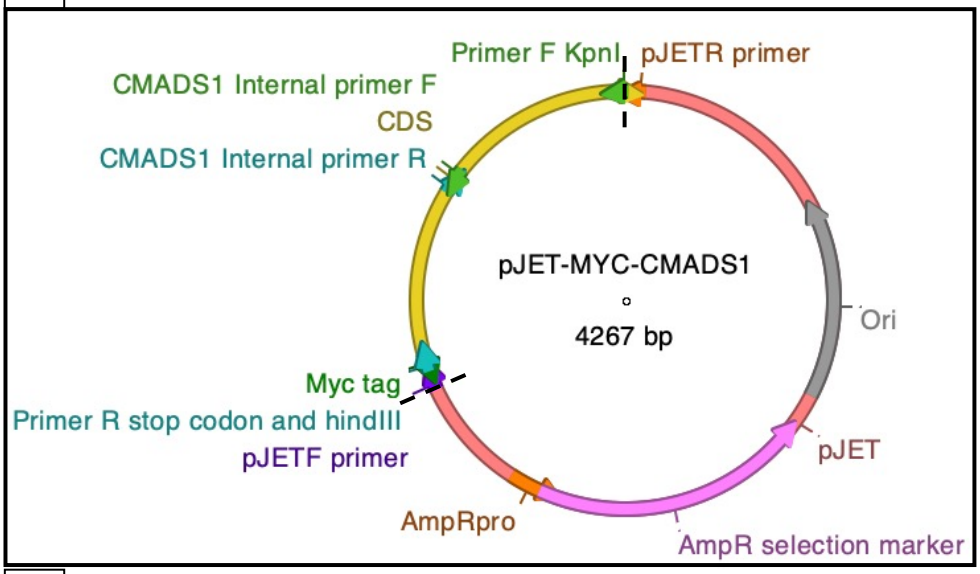
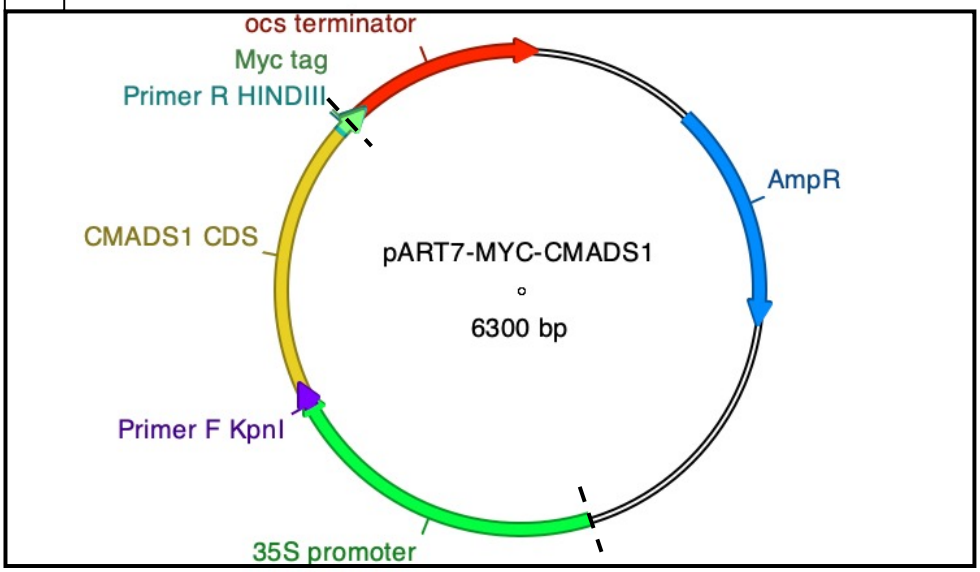

Right Border
Spec resistance gene
35S promoter
Primer F with KpnI
CMADS1 CDS
pBOMBER-35S
-CMADS1
14917 bp
Primer R with HINDIII
ocs terminator
Hyg CDS
Hyg_locus
LB
nos terminator
59bp del
Left border
C insertion
Left border

Figure 2.2: Vector construction of *35S::CMADS1* in pBOMBER for use in *CMADS1* overexpression transformations in *Ceratopteris*. Panel A shows pJET with the insertion of *CMADS1*, panel B shows the insertion of *CMADS1* into the intermediate pART7 vector using the restriction sites KpnI and HindIII. panel C shows the final construct of *35S::CMADS1* with the *35S::CMADS1* insertion cloned into the delivery vector using NotI restriction sites. Generated in ApE.

A


Primer F KpnI
pJETR primer
CMADS1 Internal primer F
CDS
CMADS1 Internal primer R
pJET-MYC-CMADS1
4267 bp
Ori
Myc tag
Primer R stop codon and hindIII
pJETF primer
pJET
AmpRpro
AmpR selection marker

B


ocs terminator
Myc tag
Primer R HINDIII
AmpR
CMADS1 CDS
pART7-MYC-CMADS1
6300 bp
Primer F KpnI
35S promoter

C

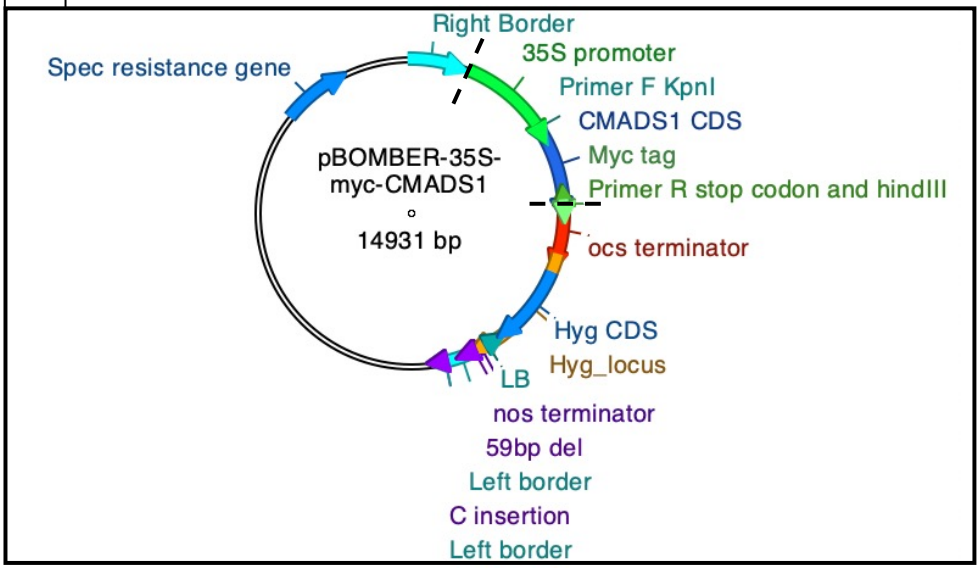

Right Border
35S promoter
Spec resistance gene
Primer F KpnI
CMADS1 CDS
Myc tag
pBOMBER-35S-
myc-CMADS1
14931 bp
Primer R stop codon and hindIII
ocs terminator
Hyg CDS
Hyg_locus
LB
nos terminator
59bp del
Left border
C insertion
Left border

Figure 2.3: Vector construction of *35S::CMADS1*-myc in pBOMBER for use in *CMADS1* overexpression transformations in *Ceratopteris*. Panel A pJET with the insertion of *CMADS1*-myc tagged, panel B shows the intermediate vector pART7 that is used to donate the overexpression promotor 35S. Panel B shows the insertion of the myc tagged *CMADS1* construct using the restriction sites *KpnI* and *HindIII* into the intermediate vector pART7, panel C shows the final construct of *35S::CMADS1*-myc with the *35S::CMADS1*-myc insertion cloned into the delivery vector using NotI restriction sites. Generated in ApE.

2.6.3. Digestion

A plasmid (pART7, pBOMBER) concentration of 1.5 µg/µl was required for digestion with CutSmart buffer and the required restriction enzyme(s) for a reaction mix of total volume 40 µL. The insert fragment digestion used the entirety of the PCR product. Digestion mixtures were incubated at 37°C for 4 hours, 1 µL Quick CIP was added to the plasmid vector digest after 3 hours to prevent any self-sticking.

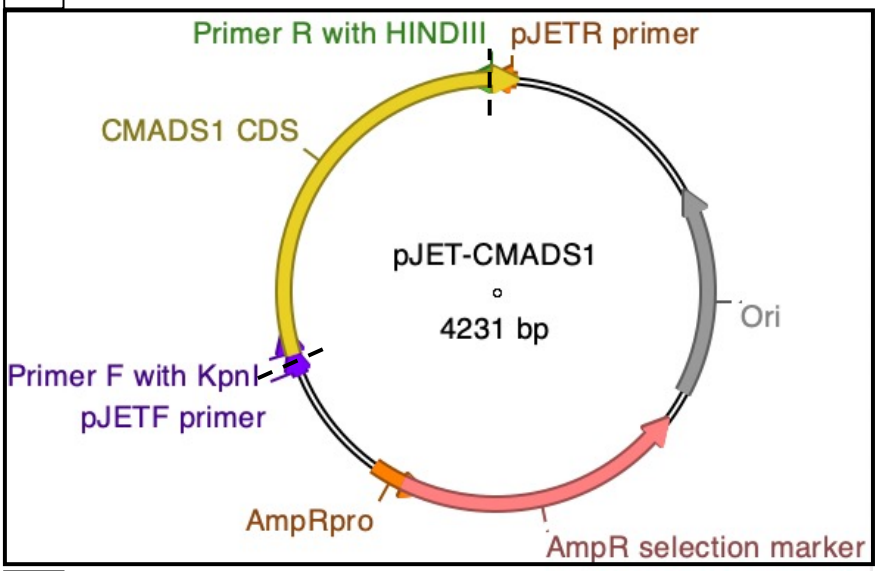
2.6.4. Purification

The digestion mixtures were ran in their entirety on a 0.8% electrophoresis gel at a constant 70V for 60 minutes for quantification for the calculations of the ligations and for gel extraction and DNA purification (as described in 2.3.4), using 50 µL wells. A final volume of 15 µL was eluted. 2 µL was ran on a 0.8% agarose electrophoresis gel for 45 minutes at a constant 70V to check purification was successful.

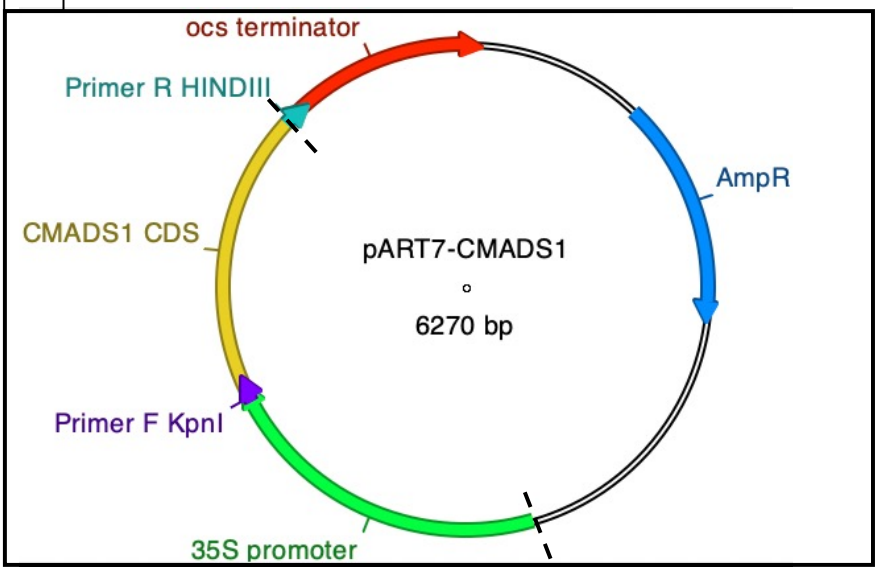
2.6.5. Ligation

The ratio of insert:vector needed was calculated using the intensity of bands on the purification electrophoresis gel as a semi-quantitative check. Multiple ratios (1:1, 3:1, 6:1) were used if ligation unsuccessful. T4 ligase buffer and T4 ligase were added along with dH_2O to a total volume of 10 µL. Vector only and no ligase controls were also performed. Multiple incubations were also used to increase the chance of a successful ligation with the reaction incubated at room temperature overnight, 37 °C for 1-2 hours, and 4 °C for 48 hours.

A


Primer R with HINDIII
pJETR primer
CMADS1 CDS
pJET-CMADS1
4231 bp
Ori
Primer F with KpnI
pJETF primer
AmpRpro
AmpR selection marker

B


ocs terminator
Primer R HINDIII
AmpR
CMADS1 CDS
pART7-CMADS1
6270 bp
Primer F KpnI
35S promoter

C

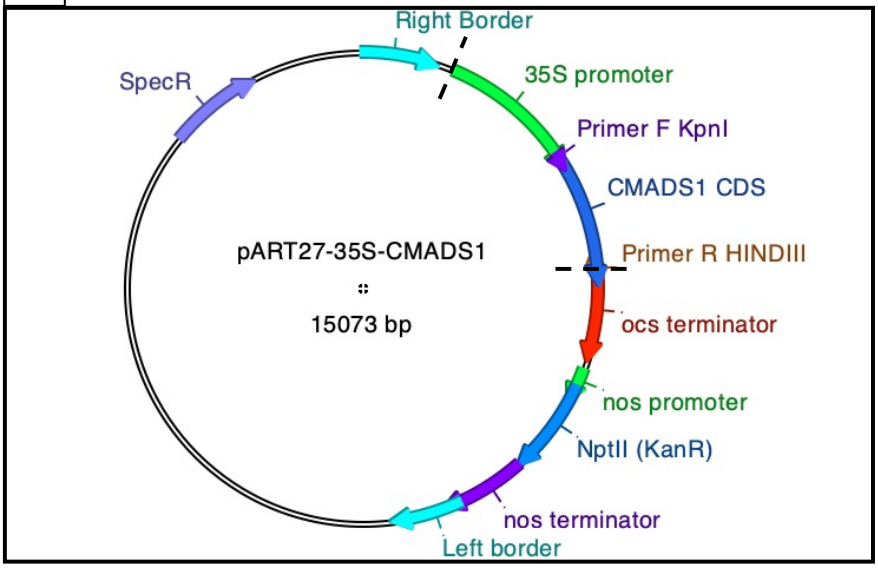

Right Border
SpecR
35S promoter
Primer F KpnI
CMADS1 CDS
pART27-35S-CMADS1
15073 bp
Primer R HINDIII
ocs terminator
nos promoter
NptII (KanR)
nos terminator
Left border

Figure 2.4: Vector construction of *35S::CMADS1* in pART27 for use in *CMADS1* overexpression transformations in *Arabidopsis.* Panel A shows pJET with *CMADS1* insertion using restriction cloning and the restriction sites *KpnI* and *HindIII*, panel D shows the insertion of CMADS1 into pART7 again using the restriction sites *KpnI* and *HindIII*, panel F shows the final construct of *35S::CMADS1* in pART27 with *35S::CMADS1* using NotI restriction sites for insertion into the delivery vector. Generated in ApE.

2.6.6. Cloning of *CMADS1* in pBOMBER for overexpression in *Ceratopteris*

The restriction cloning methods described in 2.3.5 were followed with the appropriate vectors described in Figure 2.2 and Figure 2.3 used depending on the need for a myc-tag.

2.6.7. Cloning of *CMADS1* in pANDA for RNA interference knockdown in *Ceratopteris*

Based on the phylogeny (Figure 3.3) the two genes falling into the same clade as *CMADS1* and two closest relatives were BLAST searched to check for similarities of the chosen sequence and designed primers (forward sequence: 5'-GAATGGGGCTCTGGTACA -3' reverse primer sequence: 5'-GAACAATTCCGATCCGCAG -3') no significant similarities were identified. AttB sites were then added to the primers following the protocol described by Reece-Hoyes and Walhout (2018), this protocol was followed throughout the Gateway recombinational cloning RNAi stages. With constructs described in Figure 2.5.

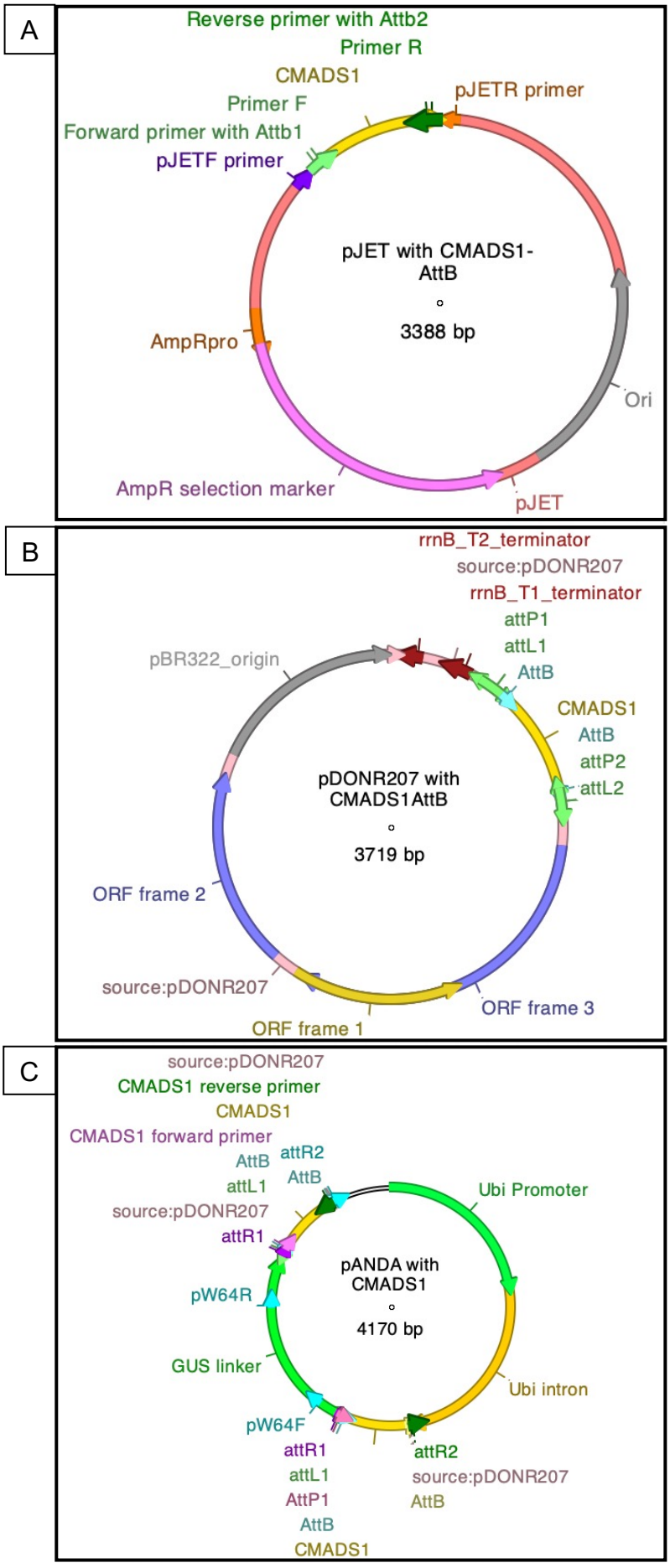

A
Reverse primer with Attb2
Primer R
CMADS1
Primer F
Forward primer with Attb1
pJETF primer
pJETR primer
pJET with CMADS1-AttB
3388 bp
AmpRpro
Ori
AmpR selection marker
pJET
B
rrnB_T2_terminator
source:pDONR207
rrnB_T1_terminator
attP1
attL1
AttB
pBR322_origin
CMADS1
AttB
attP2
attL2
pDONR207 with CMADS1AttB
3719 bp
ORF frame 2
source:pDONR207
ORF frame 1
ORF frame 3
C
source:pDONR207
CMADS1 reverse primer
CMADS1
CMADS1 forward primer
AttB
attR2
AttB
attL1
source:pDONR207
attR1
Ubi Promoter
pANDA with CMADS1
4170 bp
pW64R
GUS linker
Ubi intron
pW64F
attR1
attL1
AttP1
AttB
CMADS1
attR2
source:pDONR207
AttB

Figure 2.5: Vector construction for *CMADS1* RNAi knockdown transformation in *Ceratopteris*. Panel A shows the insertion of the chosen *CMADS1* sequence insertion as described in 2.6.3 into the vector pJET with added AttB sites. Panel B shows the intermediate vector pDONR207 with an insertion of the *CMADS1* sequence and appropriate AttB sites. Panel C shows the delivery vector pANDA, with the final construct of *CMADS1-RNAi*.

2.6.8. Preparation, initial cloning and BP reaction

Using the Phytozome database (https://phytozome-next.jgi.doe.gov) cDNA of CMADS1 annotated with each exon noted, gDNA was also annotated with exons, introns, and the 5' and 3' UTR. This was used to design primers spanning an intron to allow for differentiation between cDNA and any gDNA contamination, these designed primers had corresponding AttB sites added (see table 5). These sites are added as they will allow the expression clone to combine with the P site present on the donor vector. 1 µL – 2 µL of the cloned *CMADS1::pJET* plasmid was used in a PCR following the previously described Phusion PCR protocol, with the designed primers.

The PCR product was extracted from the agarose gel using UV for imaging and a razor blade to excise the band of the correct size. This fragment was then purified using Zymoclean Gel DNA Recovery Kit (Zymo Research) and the included protocol. To ensure no PCR errors occurred the purified PCR product was cloned into a blunt ended pJET2.1 vector and transformed into DH5alpha competent e. coli cells following the previously described lab heat shock method for transformation.

A ligation was set up for an overnight incubation at 25 °C by adding 150 ng/µL of the AttB-*CMADS1*-AttB::pJET vector was added to 150 ng/µL of pDONR207 along with the BP enzyme, and 1X TE buffer included in the Gateway cloning kit (ThermoFisher Scientific). After incubation 1 µL of proteinase K was added to the ligation and placed at 37 °C for 10 minutes. 1 µL and 5 µL of the ligation was then added to 100 µL of competent DH5 alpha *E. coli* cells following the previously described lab heat shock

method for transformation. Antibiotic selection was used to determine success, 25 µg/ml of Gentamycin in 1% Agar LB plates.

2.6.9. LR reaction

The entry clone obtained in the previous BP reaction was then used in an LR reaction with the destination vector pANDA to form the final construct for use in transgenic work. A positive control using the vector pENTR-gus alongside a negative control of empty pANDA vector were performed alongside. The steps from the previous BP reaction were followed with the BP enzyme replaced with the LR clonase included in the kit (ThermoFisher Scientific). For selection steps Kanamycin was used as this is the resistance of pANDA. Diagnostic digests were then performed by using GUS linker primers pW64F (CATGAAGATGCGGACTTACG) and pW64R (CCGAATACGGCGTGGAT).

2.6.10. Cloning of *CMADS1* and *AG* in pART27 for overexpression in *Arabidopsis*

The restriction cloning methods described in 2.3.5 were followed with the appropriate vectors described in Figure 2.4 used.

2.7. Bacterial and Yeast methods

2.7.1. Bacterial strains and growth

2.7.1.1. *Escherichia coli* DH5α strain and growth conditions

SupE44, ΔlacU169 (Φ80lacZΔM15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1.
E. coli cells were cultured on LB agar medium (Appendix 1) 90 mm petri dishes or in LB liquid cultures (Appendix 1). Media was inoculated with 100 µg/mL carbenicillin or 100 µg/mL spectinomycin or 25 µg/mL gentamycin or 50 µg/mL Kanamycin, dependent on the required selection. LB agar plates were grown inverted at 37 °C overnight, LB liquid cultures were grown at 37 °C with agitation at 200 rpm. Glycerol stocks were prepared by adding 750 µl culture to 250 µl 60% glycerol.

2.7.1.2. *Agrobacterium tumefaciens* GV3101 strain and growth conditions

A disarmed *Agrobacterium* strain produced in a C58 background with a
rifampicin resistance gene, carrying a gentamycin resistance gene was used for transformation of *Arabidopsis*.

Agrobacterium cells were grown on low salt LB agar plates with 50 µg/mL Kanamycin (for plasmid selection) and 25 µg/mL Rifampicin (for *Agrobacterium* selection) overnight at 27 °C upside down

2.7.1.3. Saccharomyces cerevisiae AH109 strain and growth conditions

MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4., gal80., LYS2::Gal1UAS-GAL1TATA-HIS3, Gal2UAS-GAL2TATA-ADE2, URA3::MEL1UAS-MEL1TATA-lacZ MEL 1.

Yeast cells were grown on YPDA (20g/L bacto-peptone, 10g/L bacto-yeast extract, 20g/L dextrose, 40mg/L adenine sulfate, 20g/L bacto-agar) agar medium in 90 mm petri dishes, inverted at 30 °C for 2-3 days. Once transformed yeast were grown on SD-LT plates (Appendix 1) inverted, at 30 °C, for a further 2-3 days. At this point yeast colonies were taken for glycerol stocks by inoculating 5 mL SD-LT liquid cultures with a colony, this was then incubated at 30 °C under 200 rpm agitation for 1-2 days until culture turns cloudy. 500 µL of this culture was then added to 500 µL 50% glycerol and then stored at – 80 °C. Colonies were also taken for further selection growth on yeast drop-out media plates of SD-HLT, SD-ALT, SD-AHLT (Appendix 1) for 3-4 days inverted at 30 °C.

2.7.2. Bacterial transformation

2.7.2.1. *Escherichia coli* DH5α transformation

Heat Shock Transformation

The ligation solution was then added to thawed competent DH5 alpha *E. coli* cells and incubated on ice for 30 minutes. Heat shock was then used to ensure the cells took up the desired construct, this was achieved by placing the cells at 42 °C for 45 seconds before placing them back on ice for 2 minutes. 500 µL of SOC media was then added to the cells and the solution incubated at 37 °C whilst shaking for 2 hours. The cells in SOC were then spread on LB medium with 1% agar in 90 mm petri dishes with 100 µg/mL Carbenicillin selection. Plates were incubated at 37 °C overnight.

Inoculation

Any colonies that grew overnight on LB plates were used to inoculate 10 mL cultures of LB containing 100 µg/mL Carbenicillin. These were then incubated at 37 °C whilst shaking overnight, to amplify bacteria for plasmid extraction.

2.7.2.2. *Agrobacterium tumefaciens* GV3101 transformation

Agrobacterium cells were spread on LB agar plates (diameter = 90 mm, supplemented with the required antibiotics) and incubated overnight (~ 16 h) at 27 °C to produce bacterial lawn that covered the surface of the plate completely. Bacterial cells were washed off the plate with 4 mL ice cold 10% (v/v) sterile glycerol. Cells growing on the

surface of the plate were scraped with an inoculation loop avoiding damages of the agar medium and were suspended in the glycerol solution. Suspensions were centrifuged at 14,000 rpm (18,000 g) for 1 min at 4 °C; the supernatant was discarded. 1 mL ice cold 10% (v/v) sterile glycerol was added and the solution vortexed to resuspend the cells. This was repeated and the supernatant was removed and discarded, the bacterial pellets were resuspended in 400 µl ice cold 10% (v/v) sterile glycerol. 70–80 µl of the ice-cold suspension of electrocompetent bacterial cells was mixed with 1–3 µl plasmid DNA (1–100 ng) in a sterile centrifuge tube. This mixture was loaded into a chilled electroporation cuvette (gap = 2 mm) and placed into the cuvette holder. The electroporator was used with the following parameters: 2.5 kV, 25 µF capacitance, and 400 Ohm resistance. 1 mL SOC medium was added immediately to the electroporation cuvette and the resulting bacterial suspension was transferred into a 15 mL centrifuge tube, and the tube was incubated at 27 °C for 1 h with rotating. When the incubation was over 100 µl from each suspension of electroporated cells was spread onto LB plates which contained the required antibiotics. The plates were incubated for 2 days at 27 °C.

2.7.2.3. *Saccharomyces cerevisiae* AH109 transformation

A colony of yeast cells were suspended in 100 mL TB mixture (33 µL Lithium Acetate 1M, 67 µL PEG 3350 50%, and 0.67 µL β-mercaptoethanol). 2 µg of each vector for transformation was added and the culture incubated at a 45 ° angle, 200 rpm, 37 °C, for 45 minutes. This solution was then spread onto SD-LT plates and the appropriate drop-out media plates. Plates were incubated at 30 °C for 2-3 days.

2.8. Yeast-II-Hybrid assays

Restriction cloning was used to produce construct of the chosen genes in both bait and prey vectors (Figure 2.6). *XmaI* restriction sites were used in the restriction cloning as described in sections 2.6.2-2.6.4. To avoid reverse orientation insertions diagnostic digests were performed as described in 2.5.6. The forward primer also excluded the start codon from the amplicon, this is due to the assay relying on the vectors own promoter (Brückner *et al*., 2009). The genes were cloned into both the bait vector pGADT7, fused with the GAL4 activation domain and the prey vector pGBKT7, fused with the GAL4 DNA binding domain (Brückner *et al*., 2009). Constructs were transformed into yeast cells and grown on appropriate yeast drop-out media as described in 2.7.2.3. Colonies were picked from SD-LT plates using sterile 10 µL inoculation loops (Sigma-Aldrich) and diluted in 100 µL nuclease free dH_2O 5 µL of the colony suspension was pipetted onto the appropriate yeast drop-out media in triplicate.

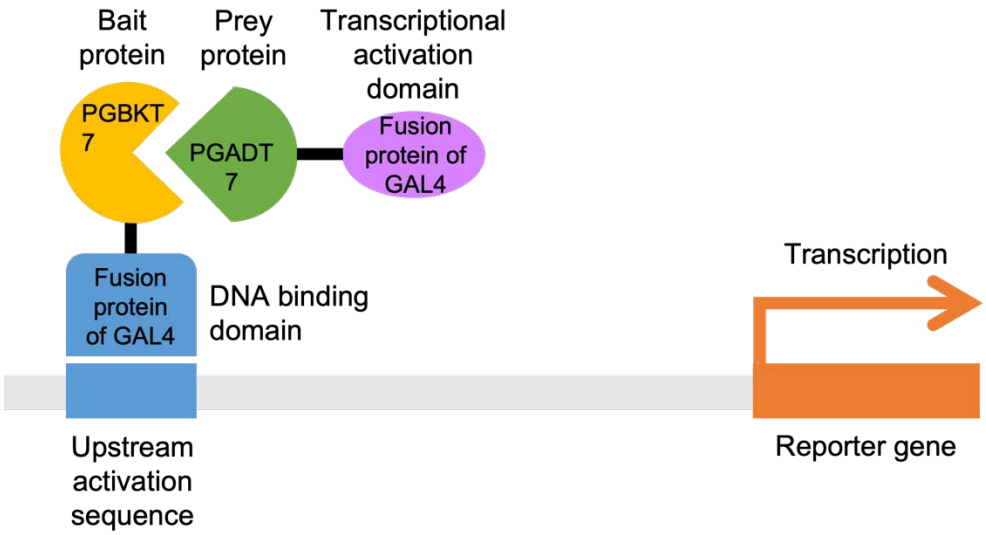


Figure 2.6: Yeat-II-Hybrid mechanism for viewing protein-protein interactions. Adapted from: How Proteins Work (Williamson, 2012).

2.9. Bioinformatics methods

2.9.1. Phylogenetics

MADS-box Type-II protein sequences were identified by searching for MADS-box proteins with both M&K domains using the Phytozome database (https://phytozome.net/) from: *Chara globularis, Coleochaete scutata, Klebsormidium nitens, Penium margaritaceum, Marchantia polymorpha, Physcomitrium patens, Anthoceros agrestis, Antheros angustus, Isoetes taiwanensis, Salaginella moallendorffii, Ceratopteris richardii, Salvinia cucullata, Azolla filiculloides, Ginkgo biloba, Pinus radiata, Picea abies, Thuja plicata, Amborella trichopoda, Arabidopsis thaliana* and, *Oryza sativa*. Presence of domains were checked individually using pfam (https://pfam.xfam.org/) as K-box domains were identified that were not initially significant during a batch search. To differentiate between MIKC* and MIKC^{c} proteins, comparison of the intervening (I) domain's length was used to differentiate between

MIKC^{C} and MIKC* proteins, with those with a shorter I-domain length being attributed to MIKC^{c}.

For comparison phylogenies in chapter 5 *Ceratopteris* and *Arabidopsis* MIKC^{c} proteins were truncated at the K-domain and M-domain as well as having full-length proteins aligned. This was done using published K-box domain sequence motifs available on phytozome.

2.9.1.1. Alignment

Peptide sequences were aligned using MAFFT (http://www.trex.uqam.ca/index.php?action=mafft) with the L-INS-i strategy. Alignments were presented using PRALINE (Simossis and Heringa, 2005).

2.9.1.2. Phylogenetic trees and model selection

The best-fitting amino acid substitution models for a maximum likelihood tree were selected using Akaike (AIC) and Bayesian (BIC) criteria in ModelTest-NG v0.1.7 (Darriba *et al*., 2019). The best-fitting models of substitution for each alignment were JTT+I+GF+F for all domains, LG+G4 for M domains and LG+I+G4 for MIK domains. Maximum likelihood trees were inferred using RAxML-NG v1.0.0 with the appropriate substitution model and 500 starting trees; 250 random and 250 parsimony trees (Kozlov *et al*., 2019). Bootstrapping was performed with an e-value cutoff of 0.01 and a maximum of 1000 bootstrap trees. Bootstraps are displayed as transfer bootstrap expectation (TBE) (Darriba, 2020).

2.9.2. Protein modelling

2.9.2.1. ChimeraX

ChimeraX (Meng *et al*., 2023) was used to present modelled secondary protein structures of both MADS-box MIKC^c proteins in *Arabidopsis* and *Ceratopteris.* The M and K domains were coloured in the structures using the predicted sequence motifs.

2.9.2.2. Alphafold2 and Google Colab

AlphaFold predictions were generated on Google Colab using ColabFold (Mirdita *et al*., 2022). *Arabidopsis* and *Ceratopteris* peptide sequences were gathered from Phytozome (https://phytozome.net/) and fed into the prediction server.

CHAPTER 3: Validating fern MADS-box genes and identifying candidate protein interactors of the published gene *CMADS1*

3.1. Introduction

3.1.1. MADS-box gene evolution

Plant proteins encoded by MADS-box genes are a family of transcription factors that have previously been categorised into the floral ABCDE model in flowering plants including representative model organisms such as *Arabidopsis* (Coen and Meyerowitz, 1991; Rijpkema *et al*., 2010). MADS-box transcription factors are responsible for regulating levels of target gene expression to generate the different identities of floral organs during development (Theissen *et al*., 2016). MADS-box genes, however, pre-date flowering plants, which first evolved around 150-190 million years ago (MYA) during the late Jurassic period (Chanderbali *et al*., 2016). MADS-box genes first evolved around 650 MYA at about the time that chlorophytes and streptophytes diverged (Becker, 2013). MADS-box proteins diverged long before the emergence of flowering plants, falling into clades separated by deep evolution (Parenicova *et al*., 2003). The MADS-box family underwent multiple evolutionary iterations before angiosperms appeared (Nam *et al*., 2003). MADS-box genes and the proteins they encode were subject to duplications and structural changes, for example going from one copy of a Type-II MIKC^{c} gene in Charophytes (Tanabe *et al*., 2005) to 107 genes of both Type I and Type II and their relative sub-categories (see introductory chapter) in angiosperms such as the *Arabidopsis*.

The wider MADS-box family is varied and diverse throughout eukaryotes (Thangavel and Nayar, 2018). MADS-box genes have been categorised in most Eukaryota with plant MADS-box genes not only containing the highly conserved 60 amino acid M

domain but also possessing the MIKC domain structure (Figure 1.6). The diversification of the MADS-box gene family is largely due to gene duplication and with the colonisation of land by green plants the family expanded vastly (Krogan and Ashton, 2000). Despite knowledge of MADS-box gene structure, the function of MADS-box genes outside of flowering plants remains largely unknown.

Previous studies in bryophytes suggested a broad function for MADS-box genes in both the gametophyte and sporophyte generations and in early plant development (Singer *et al*., 2007). Studies into the MIKC* sub-category suggest these MADS-box proteins are responsible for gametophyte development within bryophytes, whilst this function is not conserved in early vascular plants such as lycophytes (Kwantes *et al*., 2011). Despite research into the MADS-box genes present in ferns, their functions remain unknown. It has been hypothesised that fern MADS-box genes perform a ubiquitous function as those identified are largely present throughout all developmental stages of ferns (Ng and Yanofsky, 2001). A more specialised function in regulation of cell division during early organ development and an unknown function in differentiated vasculature has also been suggested (Hasebe *et al*., 1998). Understanding MADS-box genes' functions prior to the emergence of flowering plants is what this project is seeking to uncover. By looking at a plant from the sister group of spermatophytes (seed plants), the monilophytes, a valid comparison of gene function can be determined.

3.1.2. Fern MADS-box genes.

A model organism within the Monilophyte plant group, *Ceratopteris richardii,* had its genome sequence first published in 2019 (Marchant *et al*., 2019). As a result, the putative MADS-box genes annotated in the genome increased from the previously known six (Hasebe *et al*., 1998) to thirty-six (Marchant *et al*., 2019). However, these genes were not categorised further, nor were they demonstrated to be expressed, protein-coding genes. A further ten MADS-box genes were annotated from the first assembled transcriptome, bringing the total number of putative MADS-box genes in *Ceratopteris* to forty-six (Marchant *et al*., 2019). However, this number was then reduced with the second transcriptome published in 2021 to thirty-five *Ceratopteris* MADS-box genes in total (Marchant *et al*., 2022). The first genome assembly mainly relied upon short read sequencing with only a small percentage of sequencing performed at a chromosome level. A total of 38% of the genome was sequenced (Marchant *et al*., 2019). The second genome assembly prioritised working at the chromosome level and sequenced 93.5% of the *Ceratopteris* genome (Marchant *et al*., 2022).

It has been theorised that at least 2 MIKC Type-II MADS-box genes were present in the common ancestor of ferns and seed plants around 400 MYA (Henschel *et al*., 2002). The largest divergence of MADS-box genes has been observed within green plant lineages (Theissen *et al*., 2000), concomitant with their role in floral organ specification. However, the highly specific expression patterns of MADS-box genes relating to their functions in angiosperms (Zhong *et al*., 2024) is largely missing in non-seed plants where homogenous expression is often observed (Thangavel and Nayar,

2018). The majority of *Ceratopteris* MADS-box genes are more broadly expressed throughout developmental stages (Munster *et al*., 1997). However, some genes such as *CMADS1* have somewhat regulated expression (Hasebe *et al*., 1998). The functions of *Ceratopteris* MADS-box genes remain unknown (Gramzow and Theissen., 2010).

Previously, fern MADS-box genes have been shown to fall into three phylogenetic clades CRM1, CRM3, and CRM6 (Münster *et al*., 1997). These were identified from two different *Ceratoperis* species, *C. richardii* for CRM1 and *C. pteroides* for CRM3 and CRM6. The genes CRM1-CRM7 were all identified in *Ceratoperis* using Southern blotting; genes falling into CRM1 and CRM3 clades were shown to be expressed in both gametophyte and sporophyte stages of the fern life cycle using northern blotting, whist those present in the CRM6 clade showed a more differentiated pattern (Munster *et al*., 1997; Hasebe *et al*., 1998). Hasebe *et al*., (1998) showed by northern blot that *CMADS1* had differential expression between non-reproductive and reproductive tissue suggesting that a specific reproductive function could be present. *CMADS1* falls into the CRM6 clade making *CMADS1* a viable target when looking into reproductive functions of MADS-box genes in *Ceratopteris*. Interestingly, the CRM6 clade is more closely related to the AGAMOUS-like clade when compared with CRM3 and CRM1 groups (Hasebe *et al*, 1998).

One hypothesis is that due to their generalised expression, fern MADS-box genes hold a homeotic function in early cell development and organ specification, as well as overall plant architecture and organ development (Ng and Yanofsky, 2001). It has also previously been hypothesised that due to the similar expression of MADS-box genes

between reproductive and vegetative tissues that MADS-box genes within ferns have redundant functions (Hasebe *et al*., 1998).

However, the more specific expression pattern of *CMADS1* and its position in the CRM6 clade, sister to *AGAMOUS*-like clade, suggests that it could have reproductive functions making *CMADS1* a good choice for experimentation.

3.1.3. Aims and objectives of this chapter

The overarching aim of this chapter was to characterise the repertoire of MADS-box genes and their expression patterns in *Ceratopteris* and make a comparison with flowering plants including *Arabidopsis,* to enable identification of whether any MADS-box genes are likely to have a reproductive function. *Ceratopteris richardii* was the species used in this project and the species that will be referred to as '*Ceratopteris'* throughout.

To address this aim, the objectives of the work were to:

(i) Identify the evolutionary position of the *Ceratopteris* MADS-box genes using phylogenetic analysis.

(ii) Determine which of the previously annotated MADS box genes in *Ceratopteris* (Merchant *et al*., 2019; Merchant *et al*., 2022) are expressed using RT-PCR.

(iii) Define the tissue, and developmental stage-specific expression of MADS-box genes with a particular focus on potential reproductive functions, using RT-PCR and qPCR.

3.2. Results

3.2.1. Analysis of the *Ceratopteris* genome for full length MADS box protein-encoding genes.

To extend previous findings (Münster *et al*., 1997), a phylogenetic analysis was performed to visualise the relationships of Type-II MADS-box protein homologues between candidate model organisms from each land plant group (Figure 3.1). This was to determine any conserved relationships of MADS-box proteins between a non-seed plant *Ceratopteris* and the angiosperm *Arabidopsis'* MADS-box proteins. Species chosen included four green algal species to act as out-groups, the model liverwort *Marchantia polymorpha* (Bowman, 2022), the model moss *Physcomitrium patens* (Rensing *et al*., 2020), two hornwort species, two lycophyte species, 3 fern species including *Ceratopteris*, 4 gymnosperms, and 3 representative angiosperm species including *Arabidopsis*. To determine phylogenetic relationships, Type-II MADS box proteins were collated from each of these species.

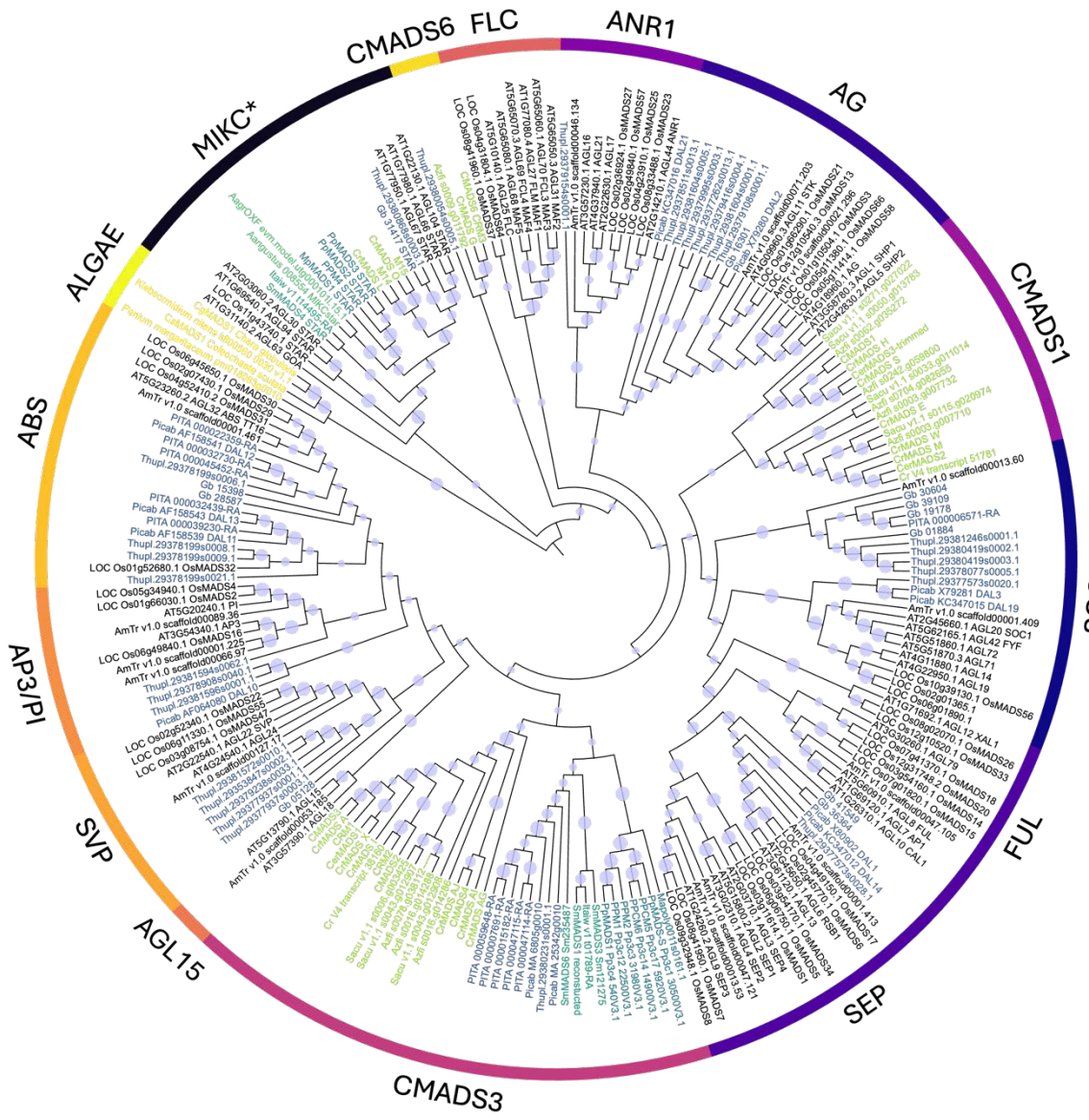

CMADS6
FLC
ANR1
AG
CMADS1
SOC
FUL
SEP
CMADS3
AGL15
SVP
AP3/PI
ABS
ALGAE
MIKC*

Figure 3.1: A maximum likelihood phylogenetic tree of MADS-box proteins from model organisms. The tree, with 100 bootstraps and rooted to four example algal species, was produced using raxml (Stamatakis, 2014). The LG model with gamma distribution was chosen by MEGAX (Kumar, Stecher, Li, Knyaz, and Tamura 2018). ITOL was used for visualisation. Yellow text: sequences from *Chara globularis, Coleochaete 76uculla, Klebsormidium nitens, Penium margaritaceum* forming the out-group of algal MADS-box proteins. Teal text: liverworts and moss sequences from *Marchantia polymorpha* and *Physcomitrium patens*. Medium green text: hornwort sequences from *Anthoceros agrestis* and *Antheros angustus*. Dark green text: lycophyte sequences from *Isoetes taiwanensis* and *Salaginella moallendorffii*. Light green text fern sequences from *Ceratopteris richardii*, *Salvinia 76ucullate*, and *Azolla filiculloides*. Blue text: gymnosperm sequences from *Ginkgo biloba, Pinus radiata, Picea abies* and, *Thuja plicata*. Black text: angiosperm sequences from *Amborella trichopoda* (a basal angiosperm), *Arabidopsis thaliana* (model dicot) and, *Oryza sativa* (model monocot). Bootstrap values are represented by lilac spheres with the smallest sphere representing 10 bootstraps and the largest 100.

Ceratopteris Type-II MIKC^c MADS-box proteins fall into the three clades previously identified as CRM1, CRM3, and CRM6 (Münster *et al*., 1997) (Figure 3.1). CMADS3 and CMADS6 identified by Hasebe *et al*. (1998) were identical to CRM1 and CRM3. The clades for this phylogenetic tree used Hasebe *et al*.'s nomenclature as that was used for *CMADS1,* the subsequently chosen gene of interest for this project. As such, the CRM6 clade was also named after *CMADS1* (Figure 3.1). The phylogeny shows that *Ceratopteris* MADS-box genes mirror *Arabidopsis* MADS-box genes by being distributed across different clades – this suggests duplication events had occurred before the divergence of ferns and spermatophytes. Due to the low resolution of the deep evolutionary branches, it is hard to establish definite patterns from this phylogeny.

3.2.2. Linking *Ceratopteris* MADS-box protein phylogeny to gene expression.

A further phylogeny was produced using the known Type-II MIKC^c MADS-box proteins in *Ceratopteris* (Marchant *et al*., 2022) alongside three MADS-box proteins present in algal species as an outgroup. The relationships observed between the MADS-box proteins within this species-specific tree corroborate those seen in the previous multi-species phylogeny. *Cr*MADS6 and CMADS3 both fall outside a clade containing *Cr*MADS_M, *Cr*MADS_W, CMADS1, *Cr*MADS_H, *Cr*MADS_S (Figure 3.2). These three clades can also be seen in the larger phylogeny (Figure 3.1), suggesting the evolutionary divergence between these gene families occurred before the divergence of the fern lineage.

Qualitative expression analysis was performed using RT-PCR of the full length CDS of these Type-II proteins using the three major stages of reproductive developmental tissue. The results are summarised in Figures 3.2. The majority of genes showed expression in all or most reproductive stages tested, with *CrMADS_G* and *CrMADS_AI* absent from mature fronds) and MIKC* genes *M14* and *CrMADS_L* showing an absence from all reproductive tissue (Figure 3.2).

As is known from the ABCDE model, MADS box proteins form quartets of proteins using both hetero- and homodimers (Parenicova *et al*., 2003). Therefore, *CMADS1* and those closely related phylogenetically were chosen for further study due to the higher likelihood of interaction and similar expression (Figure 3.2). This along with further expression analysis of MADS-box genes using RT-PCR (section 3.3) was used to choose the candidates taken forwards for quantitative expression analysis as well as protein-protein interaction analysis in Yeast-II-Hybrid assays (see chapter 5.1.2).

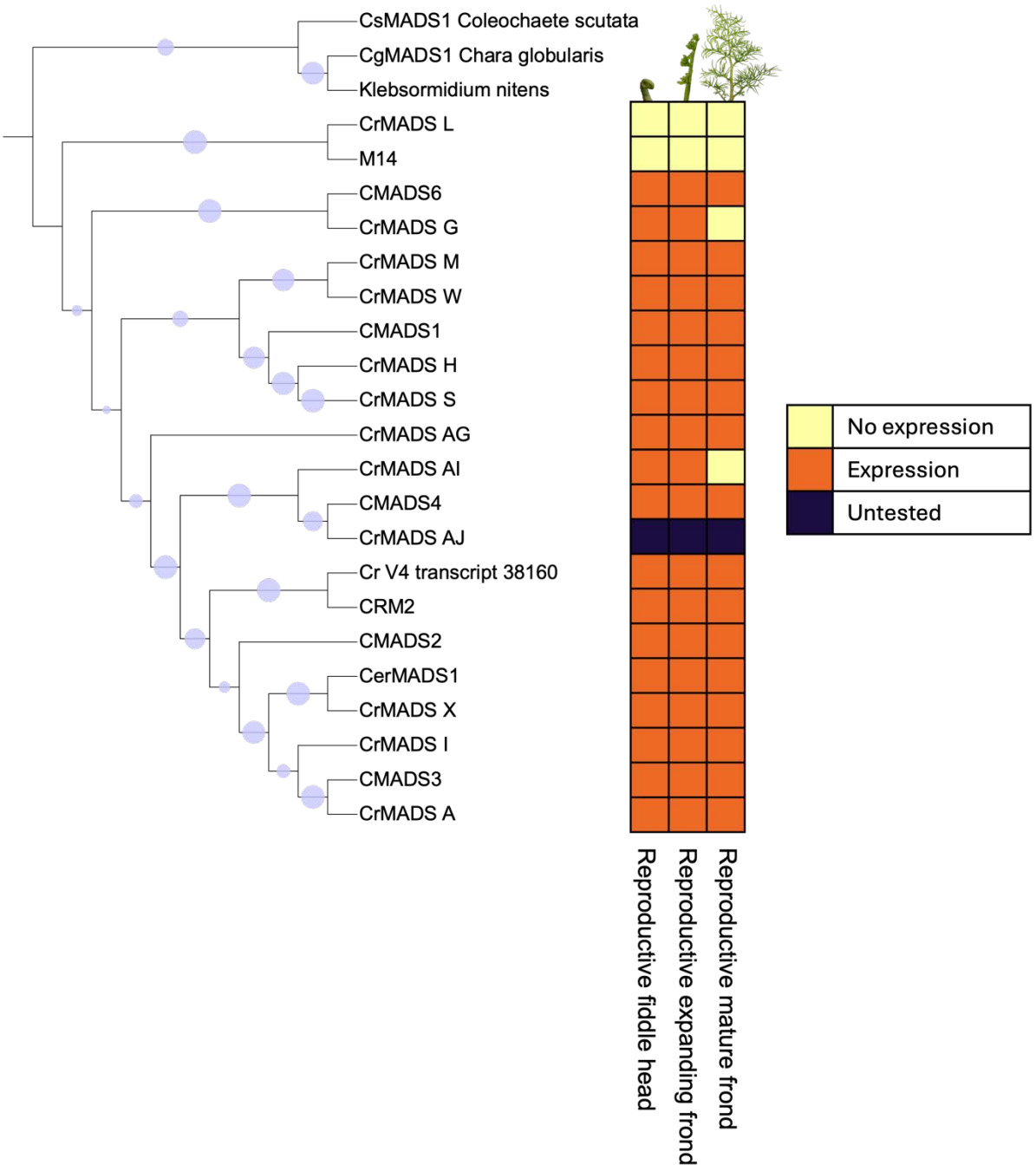


Figure 3.2: A phylogenetic tree showing relations between Type-II MIKC^{c} MADS-box proteins combined with a heat map showing the relative expression in reproductive developmental stages. Most MADS-box genes are expressed in reproductive tissue. Lilac circles represent bootstrap values with the lowest being 24 and the largest circle representing a 100 bootstrap value.

3.3. Cloning and detailed expression analysis of the MADS-box gene family in *Ceratopteris*

3.3.1. Cloning of Type-II MADS-box genes

RT-PCR to amplify and then clone and sequence full-length genes was used as a check for the validity of the genes identified initially in the first iteration of the genome assembly (Marchant *et al*., 2019) and honed when the second, better annotated, *Ceratopteris* genome (Marchant *et al*., 2022) was published. This approach also enables a qualitative check of expression patterns. The genes that showed expression throughout all reproductive stages in the initial analysis (Figure 3.4) also had further developmental stages assessed for expression via the amplification (for subsequent cloning and sequencing) of full-length genes from stage-specific cDNA.

Populations of wild-type *Ceratopteris* Hn-n were used for tissue collection of different developmental stages ranging from 5-day old gametophytes to 113-day old sporophytes tissue. This tissue was used for cDNA synthesis and cloning of candidate MADS-box genes from the different developmental stages. As a result of this cloning amino acid substitutions were found in 2 genes ("*CrMADS_R*" and "*M13*") with two further genes not being able to be cloned ("*CrMADS_Q*" and "*Cr_MADS_AJ*") suggesting the sequences mapped might not be expressed "real" genes.

The qualitative analysis of MADS-box gene expression was performed using the earlier genome assembly (Marchant *et al*., 2019) using RT-PCR (Figures 3.3 and 3.4). Due to the limited annotations associated with this genome, some Type-I genes had

their expression analysis and cloning/sequencing included. Genes that fall into this Type-I category were: *CRM1, CrMADS_D, CrMADS_E, CrMADS_F, CrMADS_R, M13, M15, M16, V4_89546* (Figure 3.3 and 3.4). Genes that were expressed in gametophytes were: *CrMADS_I, CrMADS_Q, CrMADS_R, CrMADS_D, CrMADS_G, M13, M14, M15, M16, V4_38160*, and *V4_89546* for those genes that were not Type-II further expression analysis in reproductive tissue was not undertaken. Of the Type-II MADS-box genes remaining, *CrMADS_L* and *M14* were determined to fall within the Type-II MIKC* category. Due to the long life cycle of *Ceratopteris*, gametophyte tissues were used first to amplify an many MADS-box genes as possible for cloning, sequencing and genome validation (Figure 3.3). Once the populations had grown to a sporophyte stage and a reproductive stage, further amplification, cloning, and sequencing was performed at each developmental stage to characterise genes not detected in gametophytes (Figure 3.4).

Not all genes had their expression analysed in reproductive tissue as initially the cloning was to only validate the 2019 genome (Marchant *et al.,* 2019), with the publication of the Marchant *et al.,* (2022) genome prioritisation was placed elsewhere in the project, a point of contention. One sample of cDNA was used as a positive control, this was chosen by selecting the cDNA sample that had the highest RNA concentration, to allow the stocks to last longer (Appendix 3). This was performed to show a working PCR reaction even for MADS-box genes that showed no expression. Slight differences in intensity and size of control bands can be explained by the method of running an agarose gel with the loading dye and gel red being placed into each well individually rather than in the gel itself, this was a point of contention.

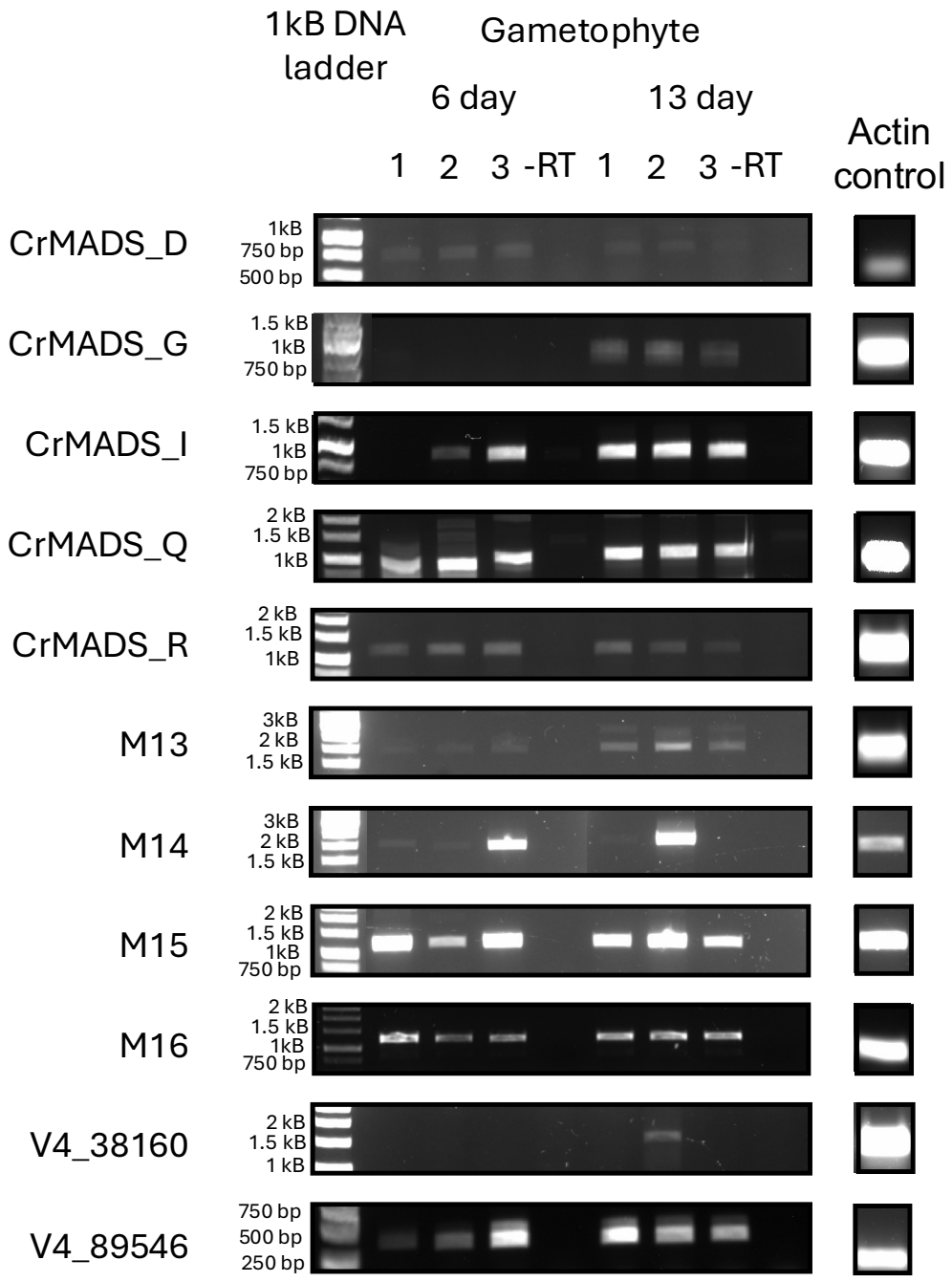


Figure 3.3: Expression patterns of Type-I and Type-II MIKC* MADS-box genes in *Ceratopteris* gametophytes. DNA electrophoresis gels showing expression of MADS-box genes in 6-day (immature) and 13-day (mature) gametophytes. 3 biological repeats are shown with negative controls lacking reverse transcriptase and a positive control using Actin.

The presence of double bands could be due to splice variants.

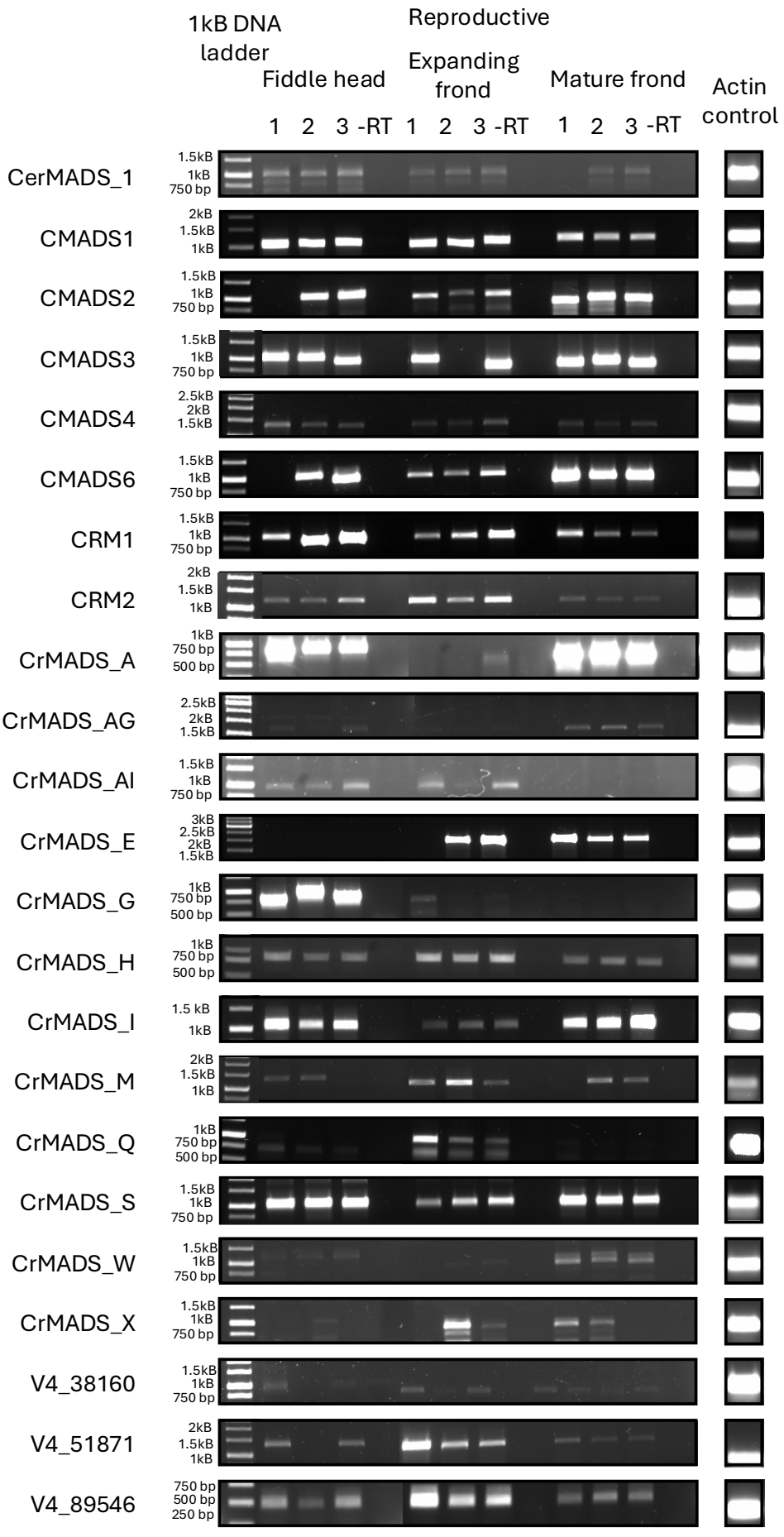

1kB DNA ladder
Reproductive
Fiddle head
Expanding frond
Mature frond
Actin control
1 2 3 -RT 1 2 3 -RT 1 2 3 -RT
CerMADS_1
CMADS1
CMADS2
CMADS3
CMADS4
CMADS6
CRM1
CRM2
CrMADS_A
CrMADS_AG
CrMADS_AI
CrMADS_E
CrMADS_G
CrMADS_H
CrMADS_I
CrMADS_M
CrMADS_Q
CrMADS_S
CrMADS_W
CrMADS_X
V4_38160
V4_51871
V4_89546

Figure 3.4: Expression of Type-II MIKC^c MADS-box genes in *Ceratopteris* reproductive tissue. RT-PCR of MADS-box genes in *Ceratopteris* 130-day reproductive fiddle head tissue, expanding frond tissue, and mature frond tissue. 3 biological repeats are shown with negative controls lacking reverse transcriptase and a positive control using *Actin*. Both controls used the RNA sample with the highest concentration from each developmental stage.

Once reproductive tissue was available for amplification, cloning, and sequencing of *Ceratopteris* MADS-box genes not all genes were tested for expression in gametophyte stages as this was done for cloning and genome validation rather than overall expression analysis. Some genes, namely *CrMADS_AJ, CrMADS_C, CrMADS_L, CrMADS_N, CrMADS_O, NCRDS_015287* showed no detectable expression in gametophyte or sporophyte tissue (Table 1).

Table 1: Genes that had RT-PCRs performed but no signals could be detected or full-length genes cloned in more than one experiment. Gametophyte tissue is represented by purple boxes, reproductive tissue is represented by teal boxes. It is likely that these genes may be pseudogenes or may have very low or restricted expression.

Gene name	Tissue tested
CrMADS_AJ	
CrMADS_C	
CrMADS_L	
CrMADS_N	
CrMADS_O	
NCRDS_015287	

3.3.2. Summary of *CrMADS* box gene name nomenclature.

As discussed in section 3.2.1, six MADS-box genes were initially discovered with *CRM1* and *CRM3* in *C. richardii* and *CRM2, CRM4, CRM5*, and *CRM6* in *C. pteriodes* (Münster *et al*., 1997). A further five were identified a year later, namely *CMADS1*, *CMADS2, CMADS3, CMADS4*, and *CMADS6* (Hasebe *et al*., 1998). *CMADS3* and *CMADS6* were identical to *CRM1* and *CRM3* respectively (Münster *et al*., 1997; Hasebe *et al*., 1998). For this project, the 'CMADS' nomenclature was used (Hasebe *et al*., 1998). Further MADS box genes were cloned by Kofuji and Yamaguchi (1997) and Kwantes *et al*. (2011). Two additional genes were identified from the genome sequence (Merchant *et al*., 2019; Plackett, unpublished) (Table 2). 7 genes were cloned for the first time during the course of this work, one of which (*V4_transcript_89546*) was not annotated in the genome sequence but was present in the Marchant *et al*., (2019) transcriptome. These genes are shown in Table 3.

Table 2: Previously described MADS-box genes and their alternative names. Type-I genes are represented by pale yellow, whilst Type-II genes are represented by magenta.

MADS-box old gene name	New gene name	Location	Type	Cloned
CMADS1/CrMADS_B	Ceric.29G021900	Chr29	Type II	HASEBE et al 1998
CMADS2	Ceric.33G065500	Chr33	Type II	HASEBE et al 1998
CMADS3/CrMADS_A/ CrMADS_AE	Ceric.1Z082000	Chr12	Type II	HASEBE et al 1998
CrMADS_C/CMADS4/ CrMADS_P	Ceric.12G051800	Chr12	Type II	HASEBE et al 1998
CMADS6/CRM3	Ceric.15G037400	Chr15	Type II	HASEBE et al 1998
CRM1	Ceric.12G052800	Chr12	Type-I	MUNSTER et al 1997
V4_38160	Ceric.16G027500	Chr16	Type II	MUNSTER et al 1997
CRM3/CMADS6	Ceric.15G037400	Chr15	Type II	MUNSTER et al 1997
CerMADS2/CrMADS_M/ CrMADS_N/V4_51781	Ceric.38G021500	Chr38	Type II	KOFUJI and YAMAGUCHI, 1997
CrMADS_S/CerMADS3	Ceric.05G029400	Chr05	Type II	KOFUJI and YAMAGUCHI, 1997
M13	Ceric.14G063000	Chr14	Type I	KWANTES et al 2011
M14/CrMADS_K/ CrMADS_J/CrMADS_L	Ceric.37G052700	Chr37	Type II	KWANTES et al 2011
M15/CrMADS_O	Ceric.39G001800	Chr39	Type I	KWANTES et al 2011
M16	Ceric.39G054900	Chr39	Type I	KWANTES et al 2011
CrMADS_E/LAMB2	Ceric.08G048300	Chr08	Type I	PLACKETT unpublished
CrMADS_F	Ceric.11G071800	Chr11	Type-I	PLACKETT unpublished

Table 3: Genes not previously cloned that have now been successfully cloned and sequenced. All but *V4_transcript_89546* being previously annotated in the constructed

genome. Gametophyte tissue represented by purple, reproductive tissue represented by teal.

MADS-box old gene name	New gene name	Location	Type	Tissue
CrMADS_I	Ceric.12G051600	Chr12	Type II	
CrMADS_R/NRCDS_015287	Ceric.16G005300	Chr16	Type I	
CRM2/V4_38160	Ceric.16G027500	Chr16	Type II	
CrMADS_G/MADS13	Ceric.27G057900	Chr27	Type II	
CrMADS_AI	Ceric.27G042700	Chr27	Type II	
V4_transcript_89546	Ceric.30G028000	Chr30	Type I	
CrMADS_H/CrMADS_H_5353350-truncated	Ceric.34G033600	Chr34	Type II	

CrMADS_I is a novel gene, previously only found in *C. pteridoides* (Munster *et al.,* 1997).

3.3.3. Potential reproductive function of MADS-box genes using expression analysis.

The expression of *CMADS1* and its closest relatives was further analysed by semi-quantitative RT-PCR throughout all developmental stages, namely gametophyte, vegetative, and sporophyte tissue (Figure 3.5). Close relatives were chosen due to the increased likelihood of similar expression. Being present in the same developmental tissues could increase the likelihood of functional protein-protein interactions, which is explored in Chapter 5. *Cr*MADS_H and *Cr*MADS_S fall into the *CMADS1* clade whilst *Cr*MADS_M and *Cr*MADS_W form the sister clade. Unfortunately, *Cr*MADS_W was unable to be cloned for sequencing due to low signal in the initial RT-PCR reaction. The semi-quantitative analysis showed that the chosen genes have quite similar expression patterns, with low or absent gametophyte expression, intermediate expression in vegetative tissue and strongest expression in reproductive tissue (Figure 3.5), making them viable candidates for future work.

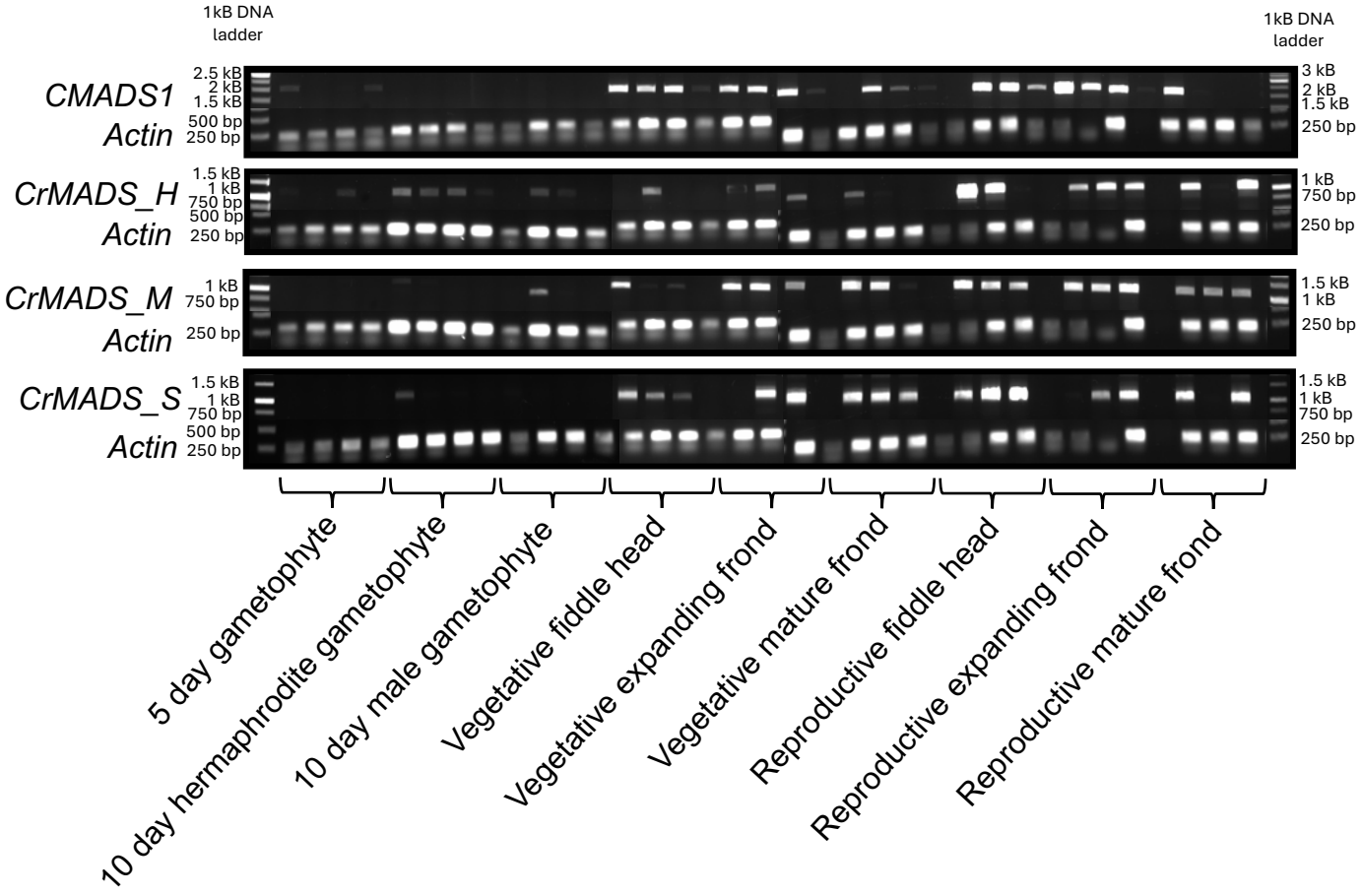


Figure 3.5: RT-PCR analysis of gene expression in all developmental stages for *CMADS1* plus three closely related genes; *CrMADS_H*, *CrMADS_M*, and *CrMADS_S*. Each reaction was run with gene-specific primers and control *Actin* primers. Two gels have been placed together for ease of viewing, with the first 17 wells being on one gel and the last 16 wells being on another. For each developmental stage, 3 biological replicates were performed as shown.

3.4. Confirmation of gene expression patterns using relative real-time quantitative PCR

In order to confirm the qualitative and semi-quantitative RT-PCR that suggested potential reproductive genes, quantitative RT-PCR (qPCR) was performed. Candidate genes were chosen using the qualitative RT-PCR expression patterns (Figure 3.3 and 3.4) coupled with the phylogeny (Figure 3.2). Genes with higher levels of expression in reproductive developmental stages than vegetative developmental stages and gametophyte stages were chosen, with those with expression patterns mirroring *CMADS1* used to determine quantitative expression. These gene fell within the *CMADS1* clade. The phylogenetic closeness could suggest an enhanced likelihood of reproductive function.

qRT-PCR analysis of *CMADS1* demonstrated statistically significant differences in expression between developmental stages (Figure 3.6). *CMADS1* shows an increase in reproductive tissues when compared to gametophytes stages and vegetative stages. This reflects what is seen in the qualitative RT-PCR (Figure 3.5), which shows little to no expression in the gametophyte stages through the absence of bands and increasing expression through the vegetative to the reproductive stages shown through different intensities of bands. Primers for qRT-PCR were designed to span an exon-exon boundary, with a Tm of 60 °C, and a size of approximately 150 bp with the cycling parameters explained in Chapter 2.5.2 followed.

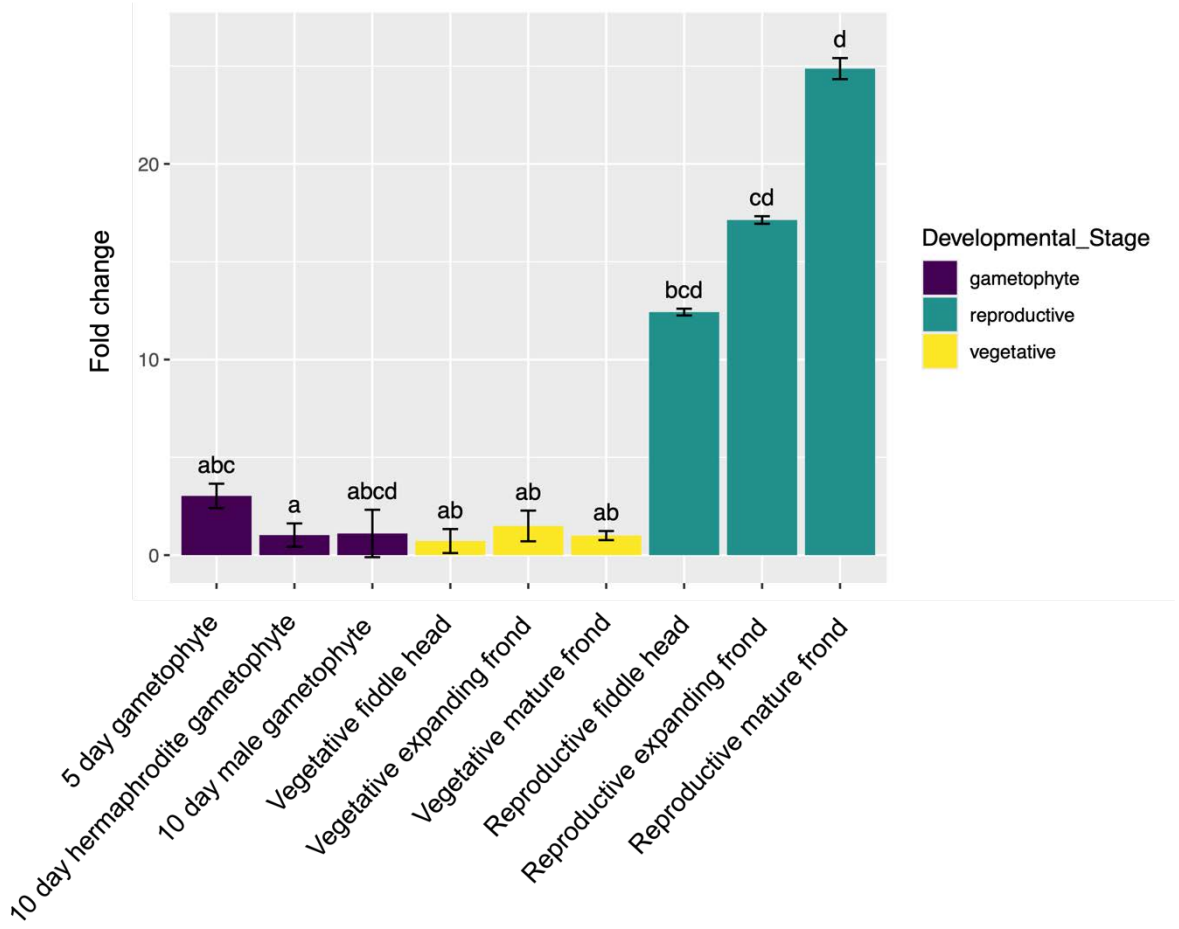


Figure 3.6: Expression of *CMADS1* across different developmental stages measured by qRT-PCR. There is an increase in expression in reproductive tissue when compared to both vegetative and gametophyte stages. The fold change was calculated using the developmental stage "Vegetative mature frond" as a reference. This was chosen as it had low expression and can offer comparison between genes. Statistical analysis via ANOVA and a post-hoc Tukey tests were used to determine significant differences between expression ($p<0.05$) of *CMADS1* at different developmental stages of *Ceratopteris* performed on the delta-CT values (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008). Letters represent statistically significant differences. Error bars represent standard error of delta-CT values of all biological replicates per developmental stage.

CrMADS_M had a similar expression pattern to that of *CMADS1* via RT-PCR, however this pattern does not translate when comparing fold changes detected by qPCR (Figure 3.7). The expression appears higher in gametophytes tissue than reproductive tissue: significant differences can be seen especially between the five-day gametophytes developmental stage and reproductive mature fronds. A trend can be seen in expression with early developmental tissue of each stage having higher expression than more mature tissue. This is consistent through the vegetative and reproductive stages with expression being higher in mature hermaphrodite gametophyte tissue than the immature gametophyte stage. This pattern of expression could suggest *CrMADS_M* holds a role in more biologically active tissue such as early developing fronds.

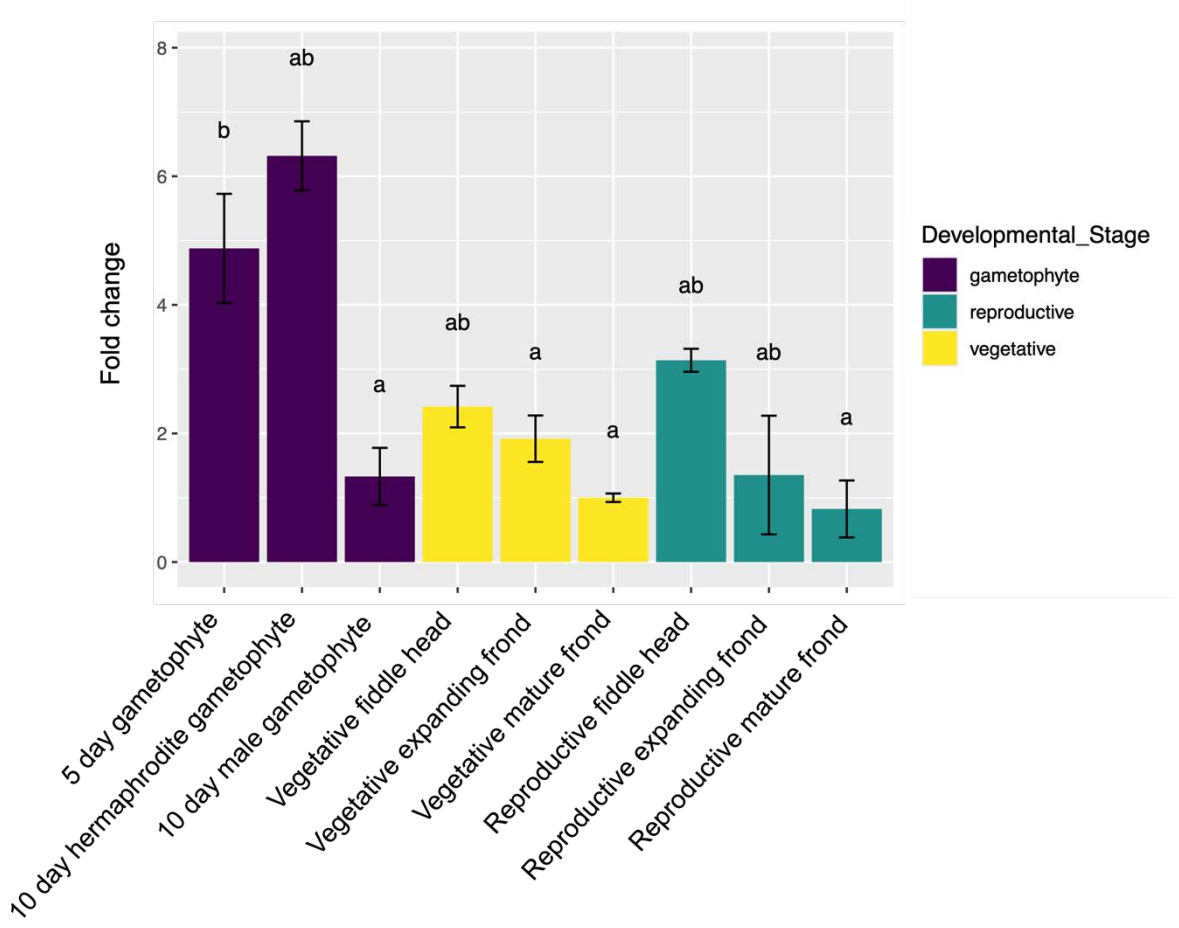


Figure 3.7: Expression of *CrMADS_M* across different developmental stages measured by qRT-PCR. There is a decrease in expression throughout each individual developmental stage being most expressed in biologically active in younger tissue. The fold change was calculated using the developmental stage “Vegetative mature frond” as a reference. This was chosen as it had low expression and can offer comparison between genes. Statistical analysis via ANOVA and a post-hoc Tukey statistical tests were used to determine significant differences between expression ($p<0.05$) of *CrMADS_M* at developmental stages of *Ceratopteris* performed on the delta-CT values (Schmittgen and Livak, 2008). Letters represent statistically significant differences. Error bars represent standard error of delta-CT values of all biological replicates per developmental stage.

CrMADS_H falls within the clade of *CMADS1* in the Type-II MADS-box protein phylogeny. It also has a similar expression pattern to *CMADS1* when observing the RT-PCR, with little expression being seen in the gametophyte stages, increased expression seen in vegetative tissue and then the most intense bands and most expressed being observed in reproductive tissue. Similarly to *CrMADS_M* this expression pattern is not reflected in the qPCR (Figure 3.8).

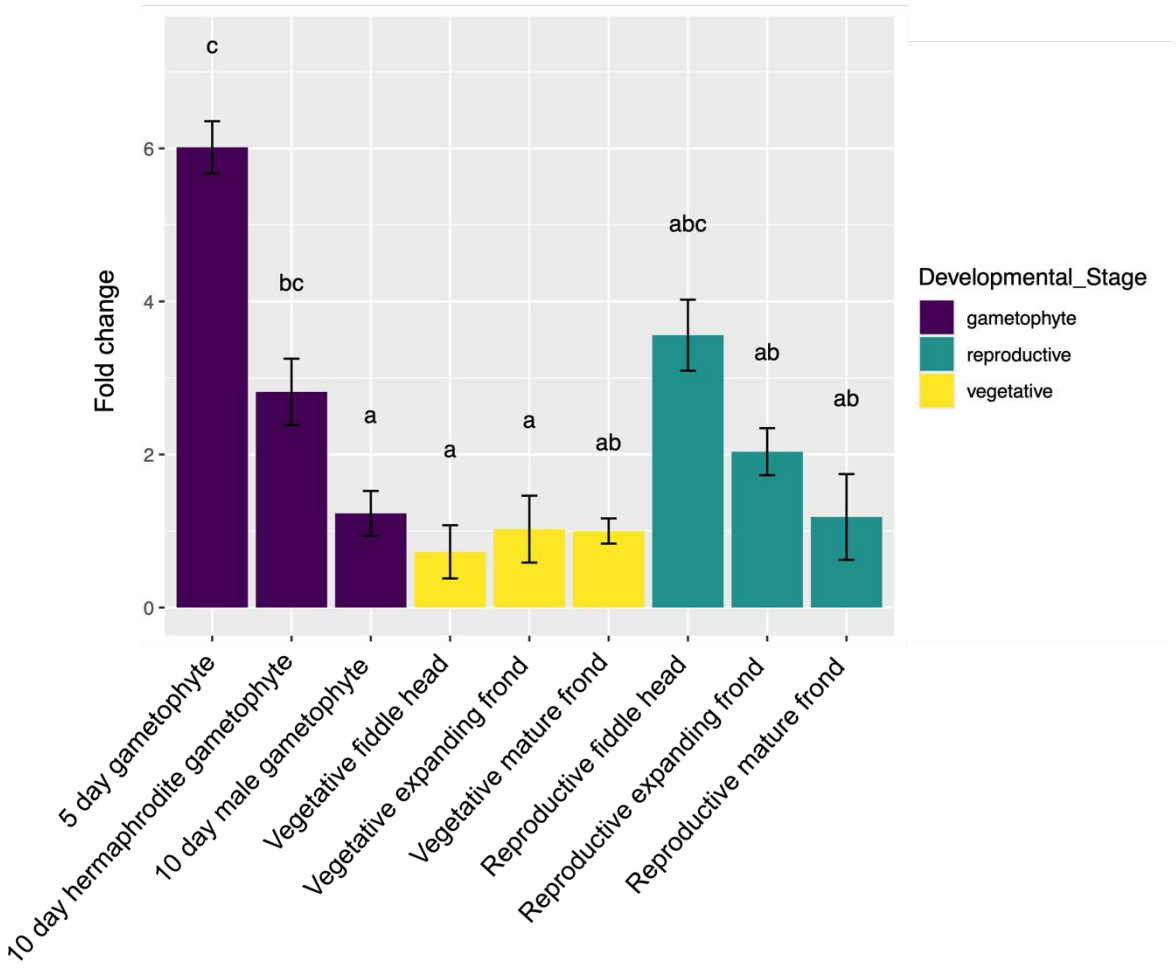


Figure 3.8: Expression of *CrMADS_H* across different developmental stages measured by qRT-PCR. There is a decrease in expression throughout each individual developmental stage being most expressed in biologically active in younger tissue. The fold change was calculated using the developmental stage “Vegetative mature frond” as a reference. This was chosen as it had low expression and can offer comparison between genes. Statistical analysis via ANOVA and a post-hoc Tukey statistical tests were used to determine significant differences between expression ($p<0.05$) of *CrMADS_H* at developmental stages of *Ceratopteris* performed on the delta-CT values (Schmittgen and Livak, 2008). Letters represent statistically significant differences. Error bars represent standard error of delta-CT values of all biological replicates per developmental stage.

CrMADS_S is also within the *CMADS1* clade in the *Ceratopteris* Type-II MADS-box protein phylogeny (Figure 3.2). Semi-quantitative expression across the different developmental stages of *Ceratopteris* shows a similar expression pattern to *CMADS1* with increased expression across reproductive tissues when compared to both gametophyte and vegetative tissues (Figure 3.2). However, the expression pattern shown under qRT-PCR has the lowest expression of *CrMADS_S* occurring in vegetative tissues and increased expression in both gametophyte and reproductive stages. Although not significantly different a decrease in expression can be seen throughout each developmental stage with the highest expression of *CrMADS_S* occurring in immature tissues (Figure 3.9).

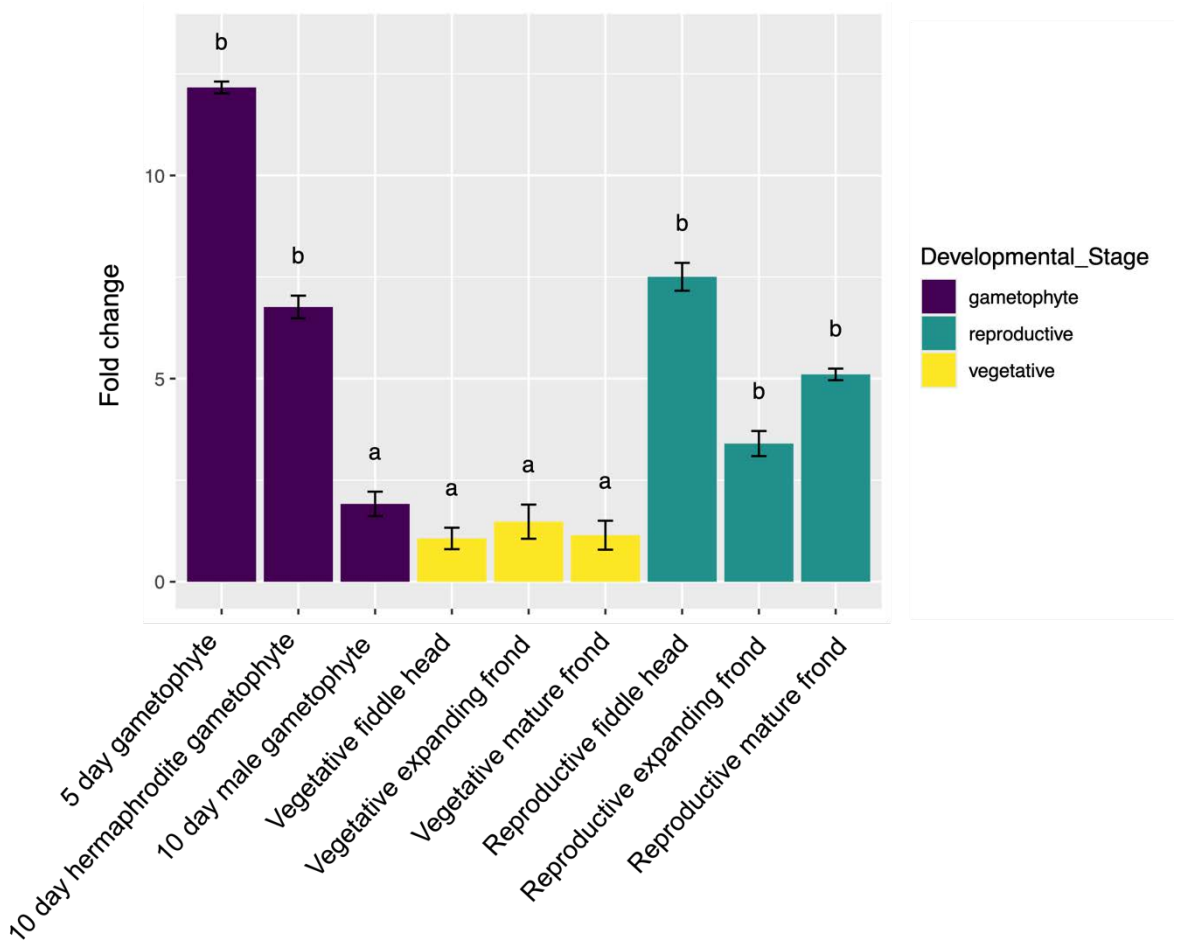


Figure 3.9: Expression of *CrMADS_S* across different developmental stages measured by qRT-PCR. There is a decrease in expression throughout each individual developmental stage being most biologically active in younger tissue. The fold change was calculated using the developmental stage "Vegetative mature frond" as a reference. This was chosen as it had low expression and can offer comparison between genes. Statistical analysis via ANOVA and a post-hoc Tukey statistical tests were used to determine significant differences between expression ($p<0.05$) of CrMADS_S at developmental stages of *Ceratopteris* performed on the delta-CT values (Schmittgen and Livak, 2008). Letters represent statistically significant differences. Error bars represent standard error of delta-CT values of all biological replicates per developmental stage.

3.5. Discussion

3.5.1. Updated MADS-box gene family in *Ceratopteris*

Initially, 15 *Ceratopteris* MADS-box genes had been identified, with 7 being cloned and assigned as *CRM1-7* (Münster *et al*., 1997). A further five were successfully cloned from the Hn-n strain of *C. richardii*, *CMADS1-4* and *CMADS6* (Hasebe *et al*.,1998). Of these, two were identified as the same as *CRM1* and *CRM3* (Hasebe *et al*., 1998). 46 putative MADS-box genes were reported in the most recent genome assembly (Merchant *et al*., 2022): This chapter reports that 33 *Ceratopteris* MADS-box genes had unique coding sequences and contained MADS-box domains, 17 contained MIKC domains of Type-II MADS-box genes whilst 14 contained the MI domains of Type-I MADS-box genes, and 2 remain uncategorised due to them being unable to be cloned. All 33 genes had their expression checked at at least one developmental stage with 18 successfully cloned and sequenced.

Of the MADS-box genes cloned in this work, only two showed differences to published sequences, suggesting the genome (Marchant *et al*., 2022) is a reliable source of MADS-box genetic information. The expression data collected suggests some *Ceratopteris* MADS-box genes show consistent expression throughout developmental stages while other genes show more specific expression patterns, this broadly reflects expression patterns seen in both plants with ancestral reproductive systems and flowering plants (Thangavel and Nayar, 2018; Zhong *et al*., 2024).

In this work, reverse transcriptase PCR (RT-PCR) was used to amplify small amounts of cDNA for use in cloning, synthesised from mRNA, allowing tissue samples to be harvested without detriment to the plant (Ghannam and Varacallo, 2018). However, PCR has known limitations such as initial changes in nucleotide sequence cause by PCR errors which then lead to errors in coding sequence (Khehra *et al*., 2023). This is a concern when cloning genes to compare to predicted genes from the published genome. This was mitigated using a proof-reading polymerase, Phusion. In this work, multiple genes did require multiple attempts to clone a gene with the same sequence as the published genome. Repeat cloning was performed for genes where initial sequencing produced potential PCR errors. After repeat cloning it was determined that no PCR errors had occurred for any of the cloning performed, rather differences to the published genome were identified, namely in the *M13* sequence and the *CrMADS_R* sequence. Both *M13* and *CrMADS_R* would be categorised as Type-I MADS-box genes with the publication of the second iteration of *Ceratopteris* genome assembly (Marchant *et al*., 2022). The updates and still relatively limited annotations of the *Ceratopteris* genome yielded multiple genes being cloned that fell into the MIKC* Type-II subfamily. These genes possess all MIK domains however they have a longer I-domain sequence containing between 2-4 more exons (Henschel *et al*., 2002).

It is also a known limitation of RT-PCR that in expressed genes incomplete cDNA copies were produced especially towards the 5’ end (Pray, 2008). During the cloning process, primers that would amplify full-length cDNA sequences were used. For those genes that showed low or no signal and as such were unable to be cloned, amplification of truncated versions could be attempted in future to further determine the presence or absence of gene expression.

3.5.2. Fern MADS-box genes have three distinct evolutionary clades mirroring the multiple clades seen in flowering plants' MADS-box genes.

MADS-box proteins are present throughout Eukaryota diversifying and multiplying as evolution progressed. MADS-box protein structures show differences as far back as the divergence of plants and animals with the structures of MADS-box proteins further diverging in plants (Thangavel and Nayar, 2018). The phylogenetic results from this chapter suggest the large number of MIKC MADS-box genes present in green land plants is dependent on multiple independent gene duplications prior to the divergence of ferns and spermatophytes. This supports and extends what was previously observed prior to the publication of the *Ceratopteris* genome and based on a limited known *Ceratopteris* MADS-box family (Münster *et al*., 1997).

A maximum likelihood phylogenetic tree was produced due to the consensus of this being the best representation of genetic relationships in deep evolution (Yang and Rannala, 2012). Although previous studies on the MADS-box family have implemented the neighbour-joining method, it is now believed that, although this method is suitable for large datasets such as the MADS-box family, the increased power of maximum likelihood is essential when inferring deep phylogenies (Ning *et al*., 2019).

Figure 3.1 is in line with previously published phylogenies of the MADS-box family with the MIKC* clade being a highly conserved clade (Qiu *et al*., 2024). The high conservation observed in the M-domain across species allows for accurate alignments

of the amino acid sequences to be determined before trimming. The model was chosen by MEGA X and the reliability reported using 100 bootstraps, although there are lower bootstrap values for deep evolutionary branches, with some reaching as low as 0, but the bootstrap values responsible for assigning clades remain within reliable parameters of a bootstrap value of 60 or greater. For the *Ceratopteris* species-specific phylogeny (Figure 3.2), maximum likelihood was again used with the model being chosen by MEGA X. This phylogenetic tree also shows that MIKC* proteins (*Cr*MADS_L and M14) form their own clade, reflecting the findings in Figure 3.1 as well as previous literature.

A recently published paper suggests that the relatedness between MIKC* and MIKC^{c} proteins could be just as distant as that previously inferred between M-type MADS-box genes (Type-I) and MIKC^{c} (Qiu *et al*., 2024). The relationships reported in this chapter support this hypothesis. Future work could potentially include Type-I MADS-box genes in a phylogeny to observe the relatedness between all three sub-categories. Comparisons between different types of phylogenetic tree would also lead to more robust data, for example comparing maximum likelihood, Bayesian, and neighbour joining could help identify highly conserved clades as well as offering comparisons of different types of relatedness. The Bayesian method is another robust method of constructing large phylogenies, similar to maximum likelihood that was used in this project; its speed could allow for more robust tests to be performed on the large MADS-box family (Douady, 2003).

3.5.3. Most MADS-box genes do not have stage-specific expression across development, with notable exceptions.

Most *Ceratopteris* MADS-box genes are similarly expressed throughout all developmental tissue stages, however those of the *CMADS1* clade (*CMADS1*, *CrMADS_S*, and *CrMADS_H*) showed increased expression in reproductive developmental stages, using semi-quantitative RT-PCR. RT-PCR semi-quantitatve expression patterns show *CMADS1, CrMADS_S, CrMADS_H,* and *CrMADS_M* to all have similar expression patterns throughout the different developmental stages of wild-type *Ceratopteris*. Low signal throughout gametophyte development suggests little to no expression of the genes, whilst weak signals present in vegetative tissues suggest some expression. The strongest signals observed in the reproductive tissues suggest high expression of all four genes in this stage of development. Having a wide range of developmental stages allowed for the observation of differing expression between early developing tissue and mature tissue.

This chapter shows discrepancies between expression patterns reported by RT-PCR and qPCR. qPCR analysis shows *CMADS1* largely reflecting the expression patterns observed in RT-PCR however the expression in gametophyte tissue, although not significantly different, is slightly higher than that seen in vegetative tissue, something that is not seen in the RT-PCR. Despite this, *CMADS1* does have s uniquely increased expression across reproductive tissues, concordant with the previous RT-PCR data as well as previous studies (Hasebe *et al*., 1998). The other three genes of interest, *CrMADS_H, CrMADS_S*, and *CrMADS_M* all show inverted expression patterns compared to their RT-PCR results, with the highest expression consistently being

viewed in gametophyte tissue and the lowest expression in vegetative tissue, with reproductive expression in between. However, what can be seen in these three genes' qPCR expression analysis is that they are more highly expressed in the early developmental stages of all types of tissue, with consistently higher expression in the gametophyte, vegetative, and reproductive early developmental tissue, and showing a pattern of decreasing as the different stages age. This is congruent to previous hypotheses that the MADS-box genes present in non-seed plants hold developmental functions of early cell differentiation and development (Münster *et al*., 1997) and may suggest that *CrMADS_H, CrMADS_S,* and *CrMADS_M* function redundantly during *Ceratopteris* development.

Relative quantification (qRT-PCR) was used to compare gene expression across different developmental tissue samples as well as to allow comparison between genes (Jalali *et al*., 2017). Discrepancies between RT-PCR and qPCR could be due to cycle number although effort was taken to keep these as close as possible with 29 being used in RT-PCR and 35 in qPCR. 384 well plates were used in the qPCR reactions although the edge effect is unlikely to be a dominating factor in the differences observed in expression between RT-PCR and qPCR experiments as the samples used are not cell cultures but rather cDNA (Mansoury *et al*., 2021). The same cDNA stocks were used for both RT-PCR and qPCR to ensure that a comparison of expression had limited variables. However, different primers were used, with RT-PCR seeking to amplify full-sequence genes and qPCR only amplifying a small section. This could be a reason for the discrepancies in expression shown between the two methods. In future work RNA-seq could be used to identify expression patterns of the

MADS-box family in *Ceratopteris,* an RNA-seq database is currently being established to provide a resource for future work with *Ceratopteris* (McCready, unpublished).

Due to the expense of qPCR only a few genes could have their expression quantified, as such a large number of Type-II MIKC^c genes present in *Ceratopteris* have only had their expression analysed qualitatively. With the discrepancies that have been observed between RT-PCR and qPCR within this chapter, it would be interesting to observe the relative quantitative expression of all *Ceratopteris* MADS-box genes.

Previously only northern blotting had been used to analyse the expression of a limited number of MADS-box genes in different tissues of *Ceratopteris* (Münster *et al*., 1997; Hasebe *et al*., 1998). In this chapter, all known Type-II MIKC^c genes have had their expression analysed semi-quantitatively through RT-PCR with *CMADS1, CrMADS_H, CrMADS_S,* and *CrMADS_M* undergoing full quantitative expression analysis. The expression of these genes has been examined throughout all different developmental stages. To build on this work more, the available transcriptome (Marchant *et al.,* 2022) could be used to produce RNA-seq data. Although all developmental stages were looked at, not all tissues were. In the future the identification of MADS-box expression within tissues such as roots could provide more insight into the function of these genes within a non-flowering plant. In addition, reporter genes using a cloned *CMADS* promoter, fused with a GUS or GFP tag would be able to show promoter activity in different tissues and different stages of development (Bui *et al*., 2015). This could add to the RT-PCR and qPCR data presented in this chapter.

3.5.4. Conclusion

In this chapter the evolutionary relationships and expression patterns of *Ceratopteris* MADS-box genes have been studied. The MADS-box proteins fall into three separate clades, extending and reflecting what was previously determined with incomplete gene lists (Münster *et al*., 1997). Although some MADS-box genes show enriched expression in reproductive tissue when observed in RT-PCR differing patterns can be observed when using qPCR. These discrepancies will need resolving in future studies. The work in this chapter provides a foundation for further functional studies. The phylogenetic closeness of *CMADS1* with *CrMADS_M*, *S*, and *H* plus their reproductive tissue expression, suggest that these proteins may function together in *Ceratopteris* reproductive development. This will be explored in the following two chapters.

CHAPTER 4: Characterising the native function of *CMADS1* in the fern *Ceratoperis richardii*

4.1. Introduction

Floral organogenesis has been well studied in model plants such as *Arabidopsis* (Parenicova, 2003). The MADS-box gene family is well categorised as containing transcription factors responsible for floral organ identity, categorised into the ABCDE model (Theißen, Melzer, and Rümpler, 2016). Loss of function mutation assays performed in angiosperms have shown function of MIKC^c MADS-box genes (Ditta *et al*., 2004). For example, a triple mutant knockout of *SEPELLATA* in Arabidopsis produced floral inflorescence of only sepals (Pelaz *et al*., 2000). As previously discussed in chapter 3, the MADS-box family pre-dates flowering plants and is composed of Type-I and Type-II subcategories with Type-II further categorised into MIKC^c and MIKC* (Figure 1.5, 1.7). This chapter focusses on MIKC^c MADS-box genes as these are the category known for floral organogenesis in angiosperms (Kaufmann, Melzer and Theißen, 2005).

Due to the high sequence and expression conservation of MIKC* type genes throughout land plants, it is likely these genes have a conserved role in gametophyte development (Zobell *et al*., 2010). The MIKC^c type genes show an increase in number reflective of the increase in complexity of the plant, for example 6 MIKC^c genes are present in the bryophyte *Physcomitrella patens* (Thangavel and Nayar, 2018) whilst 39 are present in the angiosperm *Arabidopsis* (Hileman *et al*., 2006). This conservation observed throughout extant land plants along with MADS-box gene expression in vegetative and reproductive tissue suggests a function relating to development in land plants (Thangavel and Nayar, 2018).

CMADS1 is a representative MADS-box MIKC^c gene in the model fern *Ceratopteris*. As the ferns fit into the monilophytes, the sister clade to spermatophytes, ferns could provide insight into the function of MADS-box genes prior to the advent of flowers and seeds (Münster *et al*., 1997). Previous knowledge of *CMADS1* function is largely limited to its expression pattern (Hasebe *et al*., 1998) which suggested a potential reproductive function. In contrast, other MADS-box genes known at the time showed a more homogeneous pattern of expression throughout *Ceratopteris* developmental stages and tissues, where a homeotic function likely associated with early development and cell differentiation was proposed (Münster *et al*., 1997). The expression analysis data presented in chapter 3 of this thesis supports these conclusions.

CMADS1 and those genes found in the *CRM6* clade show a similar extended protein sequence at the N terminus to *Arabidopsis* AGAMOUS; this along with increased expression in reproductive tissue could be an indicator of a reproduction-specific function (Pelayo, Yamaguchi and Ito, 2021).

4.1.1. Aims and objectives of this chapter

The aim of this chapter was to characterise the biological function of CMADS1 compared to flowering plant MADS-box function via following specific objectives:

(i) To produce transgenic *Ceratopteris* lines with both reduced and increased *CMADS1* expression,

(ii) To overexpress both *CMADS1* and its flowering plant relative *AGAMOUS* in *Arabidopsis* and investigate the effects on flowering.

CMADS1 was chosen for functional analysis using transgenic methods that have been previously developed for *Ceratopteris* (Plackett *et al*., 2015). To investigate the function of reducing *CMADS1* expression in *Ceratopteris*, RNAi was used to knock down *CMADS1* expression.

To discover the effect of increased *CMADS1* expression in *Ceratopteris*, an overexpression line was also constructed, making use of available molecular tools (Plackett *et al*., 2015). Furthermore, to investigate whether *CMADS1* could function similarly to a flowering plant MADS-box protein, constructs were used that allow for overexpression of the native *Arabidopsis AGAMOUS* for comparison, these were sourced from within the lab. Over expression of *CMADS1* and *AG* was used rather than cross-complementation of the *AG* mutant due to previous attempts at cross-complementation of *AG* being unsuccessful (Plackett *et al*., 2018).

4.2. *CMADS1* knockdown in *Ceratopteris* using RNA interference

4.2.1. Cloning of *CMADS1* gene fragments into a plant transformation vector for RNAi

A section of the *CMADS1* gene sequence was chosen for RNA interference knockdown transgenic work based on its uniqueness when compared to other MADS-box genes present in *Ceratopteris* (Agrawal *et al*., 2003) and making use of the phylogeny (Figure 3.1) generated in chapter 3. When no significant similarities were found, primers were designed to amplify the gene fragment (356 bp) and add AttB1

and AttB2 sites to be used in the initial stages of Gateway cloning (Reece-Hoyes and Walhout, 2018) (Figure 4.1).

The scheme for generating a *CMADS1-RNAi* construct in the fern transformation vector pANDA is shown in Figures 4.1, 4.2, and 4.3. A 3-step process was used whereby the *CMADS1-RNAi* fragment was first cloned into pJET to add the AttB sites required for gateway cloning (Figure 4.1), then into the intermediate plasmid pDONR207 for the recombination reaction between the pDONR207 vector and the AttB PCR product required in Gateway cloning (see Chapter 2; Figure 4.2), then finally into the fern transformation vector pANDA to generate two *CMADS1* fragments facing in opposite directions with a GUS-linker to produce double stranded RNAi in the plant (Figure 4.3). Due to the entire backbone of the delivery plasmid pANDA not being available, diagnostic digests were chosen to ensure that the constructs contain the expected insertions.

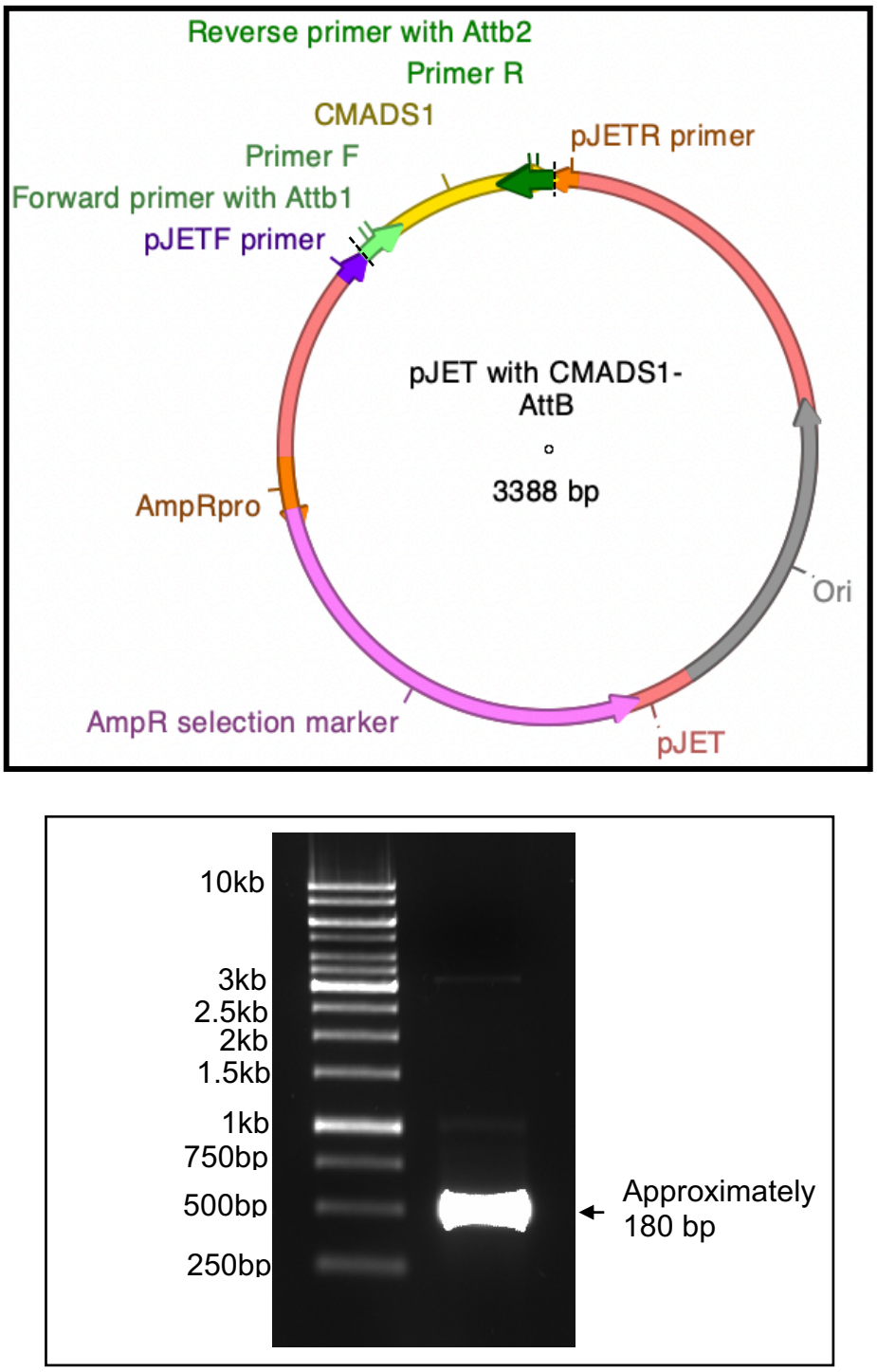
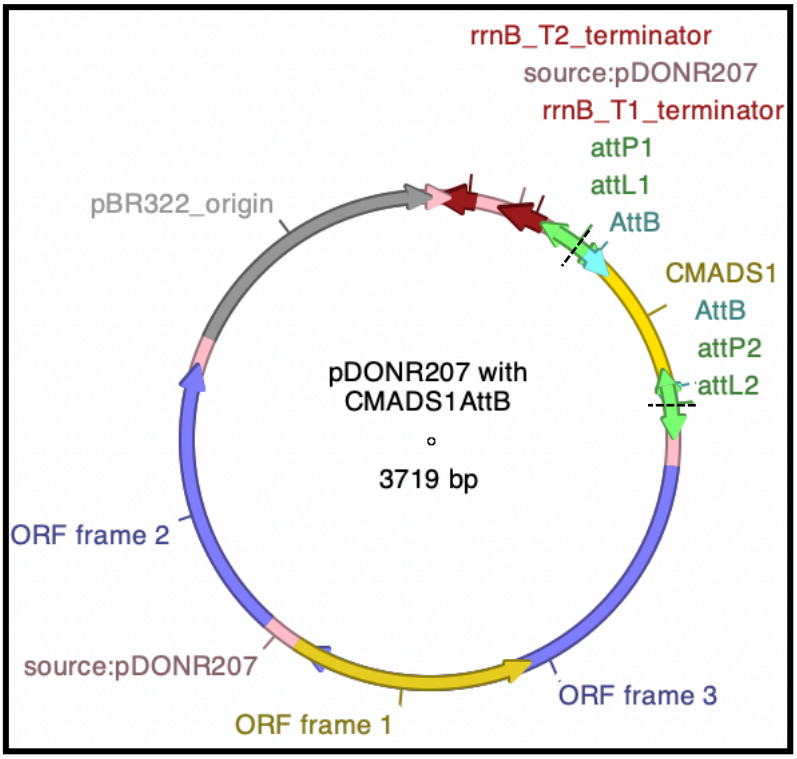


Figure 4.1: The first stage of Gateway cloning required to create the construct necessary for *CMADS1* knockdown using RNA interference. The yellow segments on the plasmids represent *CMADS1* insertions. Initial cloning of the chosen *CMADS1-RNAi* sequence by RT-PCR with AttB sites added to generate the construct CrMADS1 v1 pJET mini 1 plasmid. Gel picture shows purified PCR product of *CMADS1* with AttB sites added.


rrnB_T2_terminator
source:pDONR207
rrnB_T1_terminator
attP1
attL1
AttB
CMADS1
AttB
attP2
attL2
pBR322_origin
pDONR207 with
CMADS1AttB
3719 bp
ORF frame 2
source:pDONR207
ORF frame 1
ORF frame 3

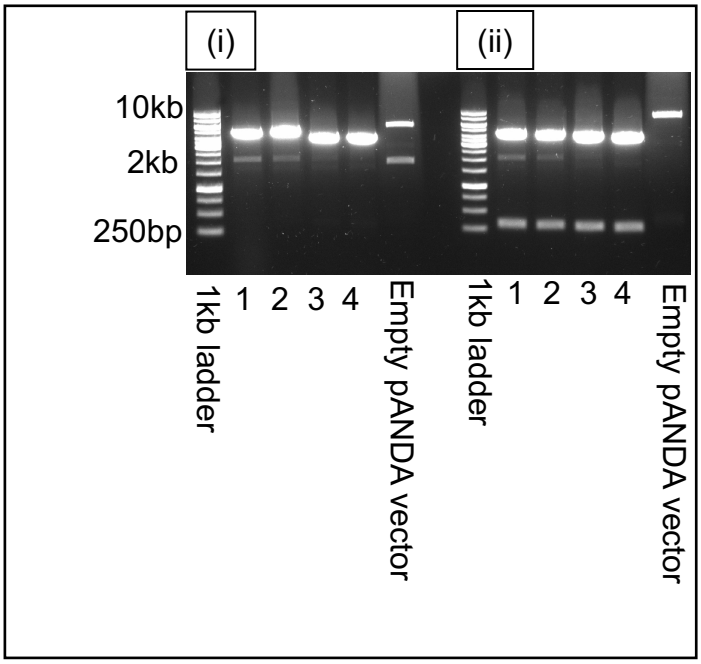
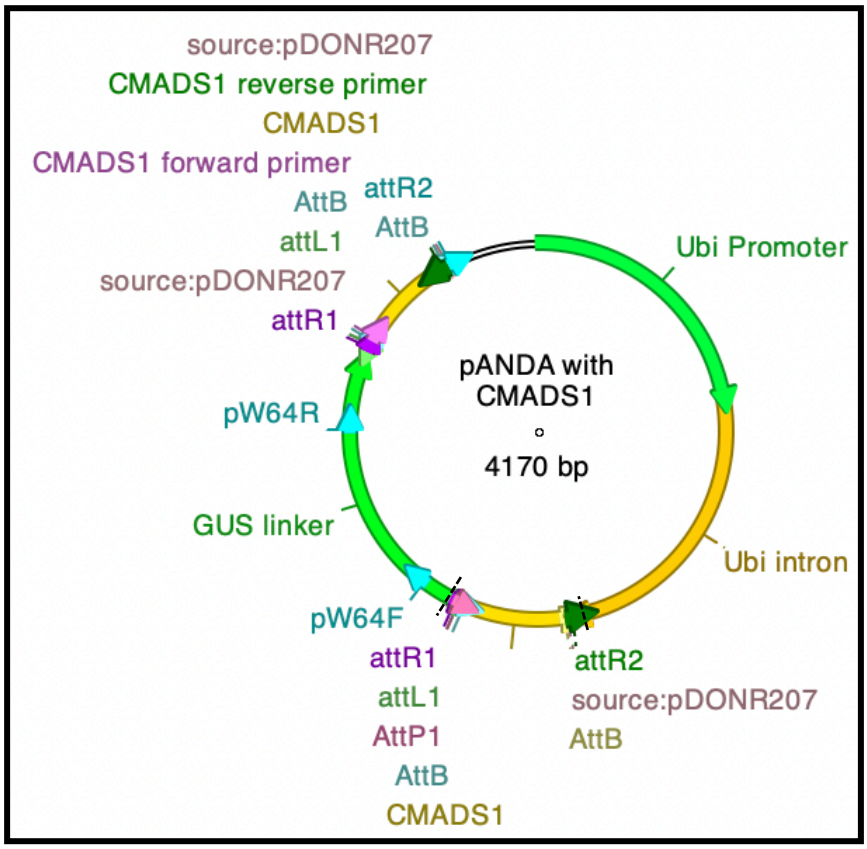

(i)
(ii)
10kb
2kb
250bp
1kb ladder
1 2 3 4
Empty pANDA vector
1kb ladder
1 2 3 4
Empty pANDA vector

Figure 4.2: The stages Gateway cloning required to create the construct necessary for *CMADS1* knockdown using RNA interference. The yellow segments on the plasmids represent CMADS1 insertions. **B)** The cloned *CMADS1* fragment in the intermediate Gateway donor vector pDONR207 and the associated correct restriction digests. (i) Diagnostic digest of pDONR207 with *CMADS1* construct with *ApaI* and *NcoI* - a successful insertion has an expected band size of 3725 bp, an untransformed vector has expected bands of size 4164 bp and 1421 bp. This is due to the CMADS1 insertion containing an *NcoI* restriction site. The digest was a double digest to increase the differences between a transformed and untransformed vector band sizes. The negative control empty vector has expected bands of sizes 4164 bp and 1421 bp. (ii) Diagnostic digest of pDONR207 with *CMADS1* construct with *NheI*, a successful insertion has expected bands of size 3459 bp and 266 bp, an unsuccessful transformation has expected bands off size 5319 bp and 266 bp. The negative control empty vector has expected band of size 5319 bp and 266 bp.


source:pDONR207
CMADS1 reverse primer
CMADS1
CMADS1 forward primer
AttB
attR2
AttB
attL1
source:pDONR207
attR1
pW64R
GUS linker
pW64F
attR1
attL1
AttP1
AttB
CMADS1
pANDA with
CMADS1
4170 bp
Ubi Promoter
Ubi intron
attR2
source:pDONR207
AttB

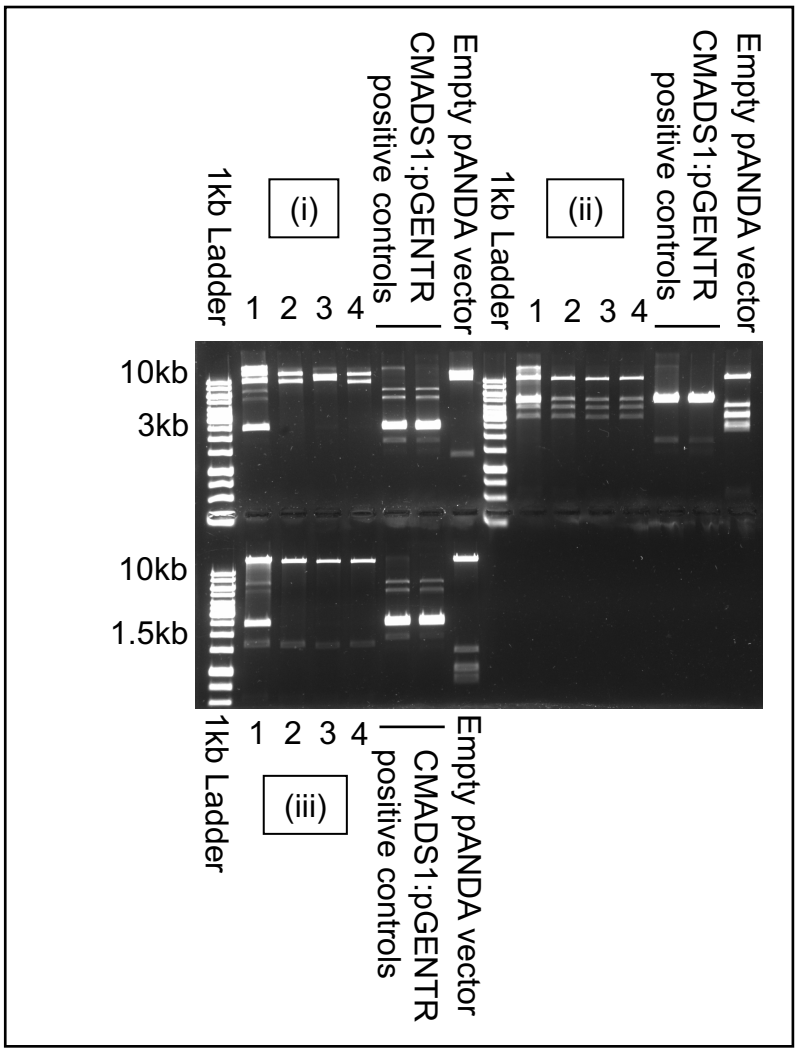

1kb Ladder
1 2 3 4
(i)
positive controls
CMADS1:pGENTR
Empty pANDA vector
1kb Ladder
1 2 3 4
(ii)
positive controls
CMADS1:pGENTR
Empty pANDA vector
10kb
3kb
10kb
1.5kb
1kb Ladder
1 2 3 4
(iii)
positive controls
CMADS1:pGENTR
Empty pANDA vector

Figure 4.3: The stages Gateway cloning required to create the construct necessary for *CMADS1* knockdown using RNA interference. The yellow segments on the plasmids represent *CMADS1* insertions. Final construct used for RNAi, named pANDA::*CMADS1*-RNAi. (i) Diagnostic digest of pANDA with *CMADS1* construct with *NotI*. Successful insertion of *CMADS1* into pANDA gave expected bands of size 3.3 kb and 764 bp. Empty pANDA vector digest gave expected bands at sizes 3.5 kb, 1.9 kb, 1 kb, and 168 bp. (ii) Diagnostic digest of pANDA with *CMADS1* construct with *NcoI*. Successful insertion of *CMADS1* into pANDA gave expected bands at sizes: 3.1 kb, and 964 bp. Empty pANDA vector expected bands at sizes: 2.5 kb, 2 kb, 1.2 kb, and 964 bp. (iii) Diagnostic digest of pANDA with *CMADS1* construct with *BamHI*. Successful insertion of *CMADS1* into pANDA gave bands at sizes: 19kb, 1.4kb, 532bp, and 275bp. Empty pANDA vector gave bands at sizes: 1.2 kb, 846 bp, 839 bp, 703 bp, 703 bp, 275bp, 262 bp.

4.2.2. *Ceratopteris* transformation of *CMADS1*-RNAi construct

Once the *CMADS1* RNAi and over expression constructs were prepared, callus was stimulated to produce undifferentiated cells that could then be transformed with the RNAi construct using microparticle bombardment (Plackett *et al*., 2014) (Figure 4.2).

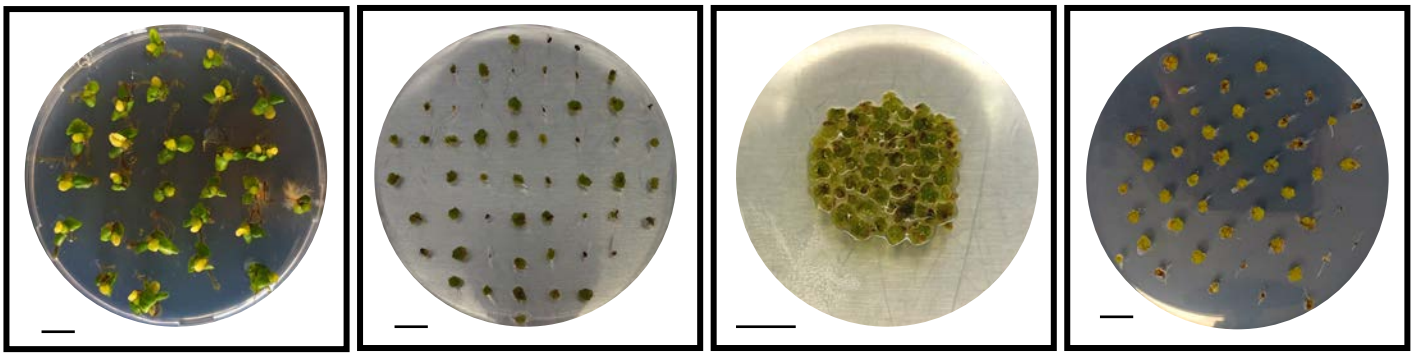
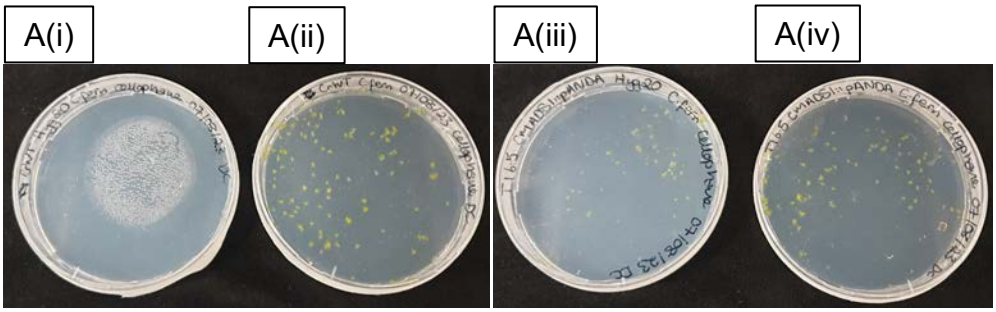


Figure 4.4: Callus generation and bombardment for transformation of *Ceratopteris*. Left to right: callus induction using the plant hormone BAP, callus separation and growth on MS-media, callus prepared for bombardment, callus on Hygromycin selection media post bombardment.

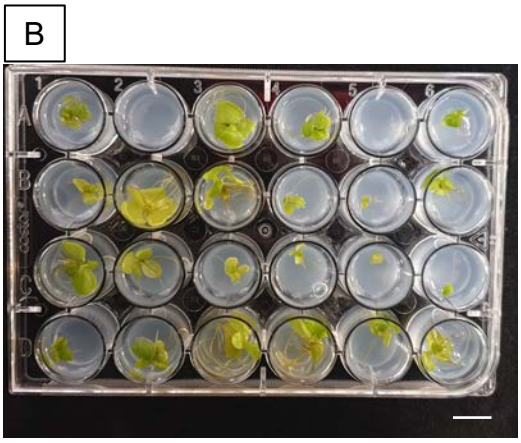
Multiple bombardments were performed for each construct. The limitations of the biolistic gun allowed for 4 DNA-coatings each time, which allowed for a different DNA-coating to be used for each plate. This meant 3 replications were performed for each bombardment. Six bombardments were performed in total for pANDA::*CMADS1-RNAi*, and six bombardments were performed in total for the empty pANDA vector controls where empty pANDA vector DNA was used to bombard callus to observe any differences in growth that could be attributed to the bombarding process and vector insertion. One replication was reserved for a no-DNA control that was bombarded to ensure no contamination when screening for transgenics (Plackett *et al*., 2014). After

transformation, bombarded callus underwent Hygromycin selection (Figure 4.5A). At this stage, resistant hermaphrodite gametophytes were isolated (Figure 4.5B), possible due to the morphological differences between hermaphrodite and male gametophytes (shown in Chapter 1), to increase the likelihood of homozygous offspring via self-fertilisation (Plackett *et al*., 2015). Hygromycin selection was also performed at the sporophyte stage (Figure 4.5C) to ensure production of stable transformants (Plackett *et al*, 2014). In the primary transformed (T0) generation, 23 transformants were taken to soil per bombardment replication. At the T1 generation, transformants underwent the same antibiotic selections and four plants were taken to soil from each line: space constraints in growth cabinets limited the number of transformants propagated.

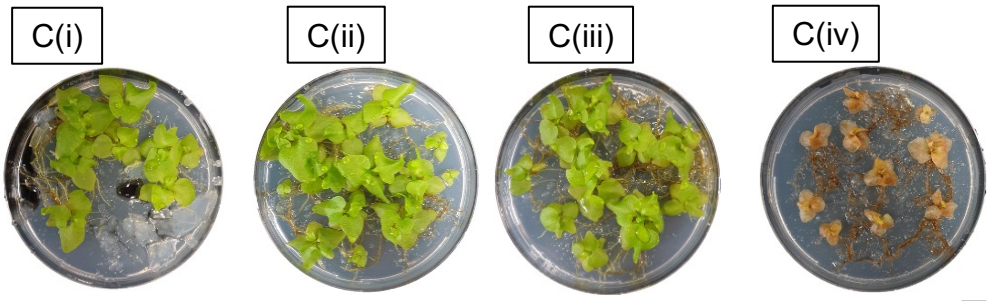
Two-week gametophytes on C. fern with Hygromycin selection and control plates media



Two-week gametophytes isolated on MS-media



Sporophytes on MS-media with Hygromycin selection



Sporophytes bombarded with *CMADS1-RNAi*

Sporophytes bombarded with a no-DNA control

Scale bars represent 1 cm to scale

Figure 4.5: Selection of T1 *Ceratopteris* transformed with *CMADS1*-RNAi vector. Panel 1 shows selection plates of T1 gametophytes using Hygromycin in C. fern medium, 1A shows Hn-n wild-type *Ceratopteris* on selection with no growth visible, 1B shows Hn-n wild-type *Ceratopteris* growth on a control plate. 1C shows growth of a transformed line on selection, 1D shows growth of transformed *Ceratopteris* on control media. Once isolated gametophytes had fertilisation induced to produce sporophyte plants on MS medium as is shown in panel 2. Once true fronds were produced, plants underwent secondary hygromycin selection of T1 C *Ceratopteris* sporophytes as is shown in panel 3. 3A-C show transformed plant growth on selection whilst 3D shows wild-type Hn-n *Ceratopteris* control on a selection plate.

4.2.3 *CMADS1* RNA interference genotyping and phenotyping of T1 lines

In order to ascertain whether RNAi of *CMADS1* led to a change in reproductive morphology or spore production, phenotyping was undertaken as described in section 4.2.3. Alongside this, to ascertain whether RNAi of *CMADS1* led to a reduction of gene expression in transgenic fern lines, quantitative PCR (qPCR) was used to genotype wild type Hn-n plants, two empty vector control lines and four *CMADS1-RNAi* lines to compare *CMADS1* expression levels (Figure 4.6). Frond length was calculated in *CMADS1-RNAi* transgenic plants and both empty vector and untransformed wild type Hn-n controls at 138-days after fertilisation - a maximum of four plants were assayed per biologically independent line (Figure 4.6). One frond was taken from each plant and measured using ImageJ (Schneider, 2012). 4 plants were grown for each line of T1 spores, as growth space was limited.

When frond length was measured in 4 plants of each of the *CMADS1-RNAi* lines, 2 empty vector lines, and a wild-type control, a Shapiro-Wilk normality test gave a p-value of 0.358. As this is above the 0.05 threshold, the data for each line can be assumed normal and parametric statistical tests performed. To compare frond length between the lines, an ANOVA was performed which gave a highly significant p value of <2e-16. A post-hoc Tukey test showed a significant difference between wildtype and *CMADS1-RNAi* line 1.3, and another significant difference was found between wildtype and the empty pANDA control line (Figure 4.6).

When number of branches per frond was measured in 4 plants of each of the *CMADS1-RNAi* lines, 2 empty vector lines, and a wild-type control, a Shapiro-Wilk normality test gave a p-value of 0.1093. As this is above the 0.05 threshold the data can be assumed normal and parametric statistical tests performed. To compare number of branches per frond between the lines, an ANOVA was performed which gave a highly significant p value of 0.0023. A post-hoc Tukey test showed a significant difference between wildtype and *CMADS1-RNAi* line 6.1. Other significant differences reported by the ANOVA were determined to lie between the various RNAi transgenic lines to each other (Figure 4.6).

When number of pinnae per frond were measured in 4 plants of each of the *CMADS1-RNAi* transgenic lines, 2 empty vector lines, and wild-type controls, a Shapiro-Wilk normality test gave a p-value of 0.1816; as this is above the 0.05 threshold the data can be assumed normal and parametric statistical tests performed. To compare number of pinnae per frond between lines, an ANOVA was performed which gave a highly significant p value of 0.0022. A post-hoc Tukey test showed no significant difference between wildtype and the transgenic lines, but significant differences reported by the ANOVA were determined to lie between the various *CMADS1-RNAi* lines (Figure 4.6).

When number of sporangia per mm of pinna was calculated in 4 plants of each of the *CMADS1-RNAi* transgenic lines, 2 empty vector lines, and wild-type controls, 3 pinnae per plant were used for measurements. This was then used to calculate the average number of sporangia per mm of pinnae. 4 plants of each of the *CMADS1-RNAi* transgenic lines, 2 empty vector lines, and wild-type controls, a Shapiro-Wilk normality

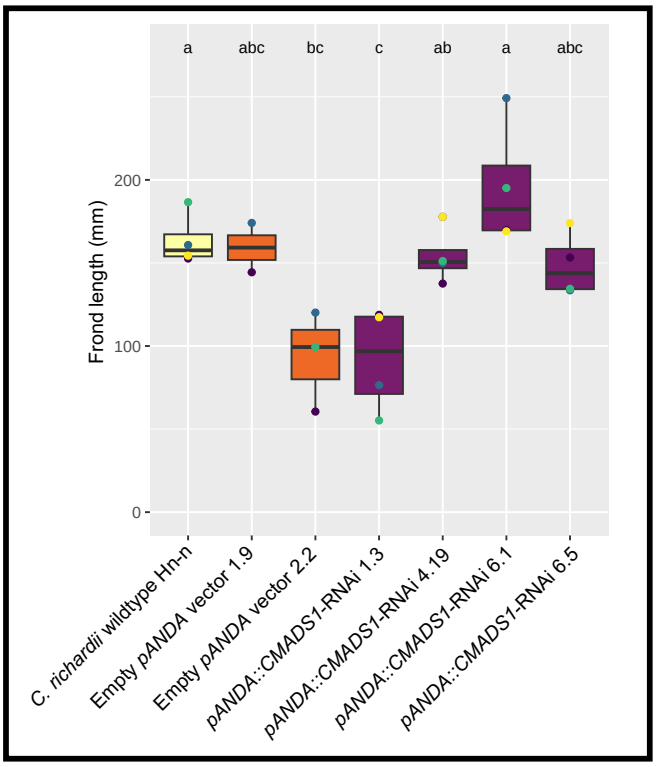
test was performed giving a p value less than the threshold of 0.05 (8.621e-06). This determines that the data is not of normal distribution as such the non-parametric test of a Wilcoxon Rank Sum test was performed between the *Ceratopteris* wild-type control and each transgenic line. A significant difference was observed between empty pANDA vector control line 2.2 and *CMADS1-RNAi* line 1.3 (4.7).

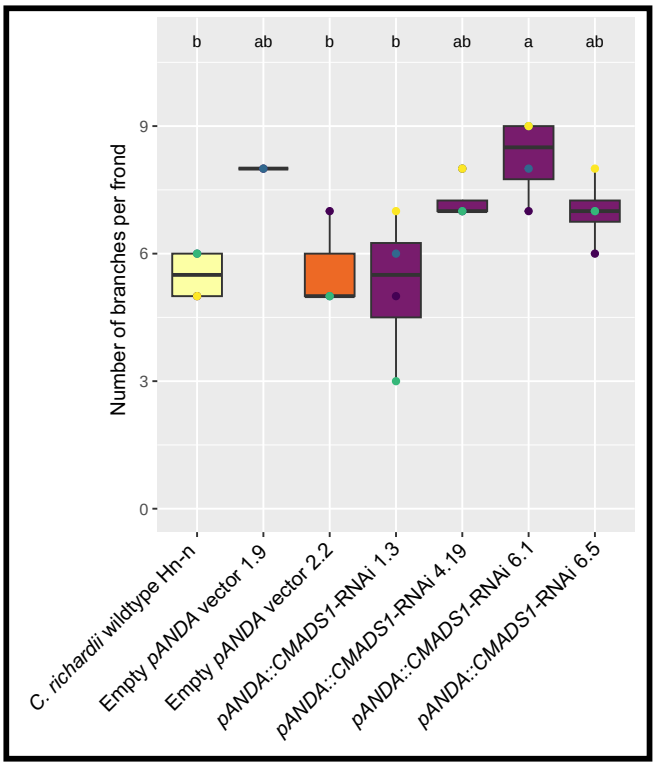
For genotyping, a reproductive fiddle head was collected at 138-days from each plant in each line for RNA extraction and use in qPCR. Fronds were harvested at the same time stage of the fronds used for phenotyping to offer better comparison. qPCR was used to determine quantitative levels of *CMADS1* expression. By 138-days senescing had taken hold, as such not all RNA samples gave reliable cDNA stocks and some data points had to be disregarded due to large variability between technical replicates. A total of four technical replicates were performed to allow for outliers to be identified.

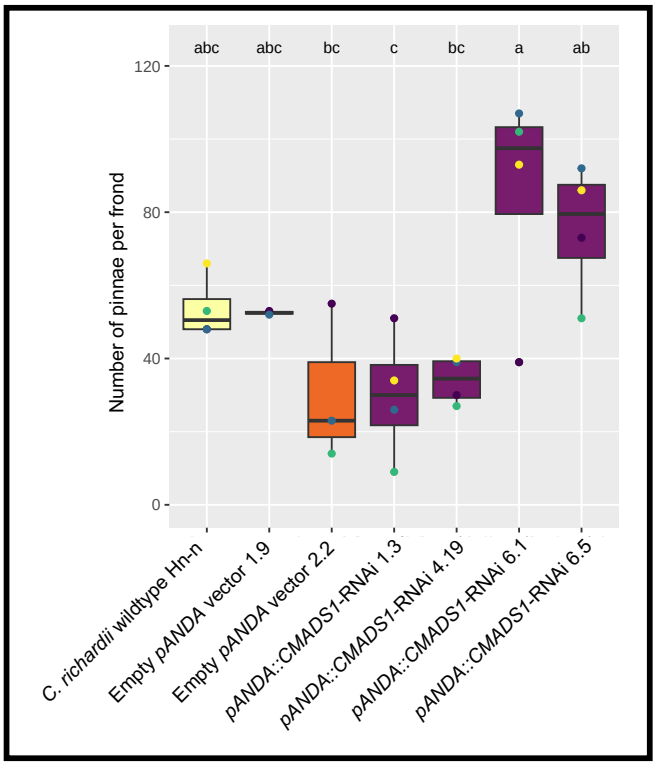
No significant differences in *CMADS1* expression levels were seen between lines, and no consistent differences between the level of *CMADS1* expression in wild type/pANDA controls and *CMADS1-RNAi* lines were observed (Figure 4.6 and 4.7). Moreover, no consistent correspondence was seen between the level of *CMADS1* expression measured by qRT-PCR and any of the recorded phenotypes (Figure 4.6 and 4.7). Taken together, these data suggest that the RNAi knockdown of *CMADS1* may have been unsuccessful, and that the phenotypic variation seen between lines is independent of *CMADS1* expression level.

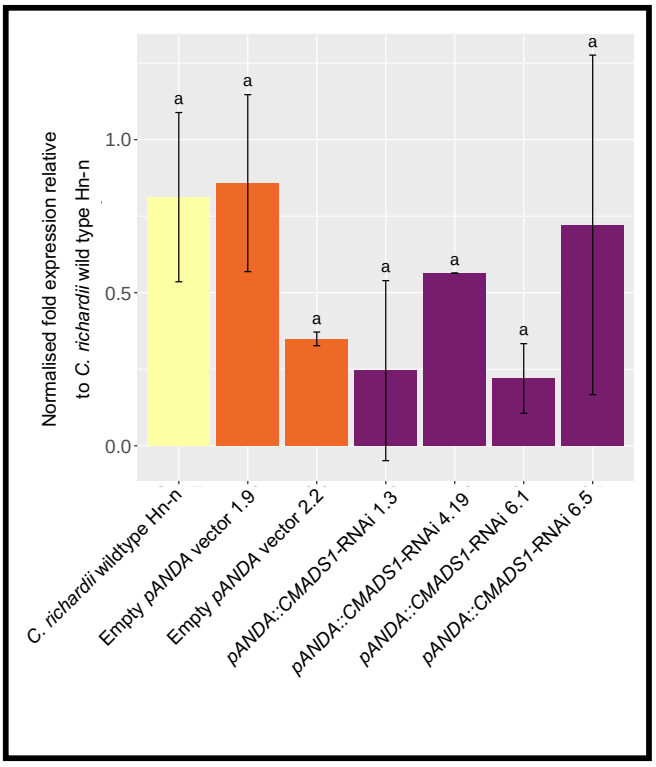
The empty vector controls also show significant differences in *CMADS1* expression when none are expected, suggesting a variation in data independent from the

knockdown of *CMADS1*. The small number of replicates is a limiting factor in the validity of these results with constraints on growing space and the necessity of preserving the plants for further spore collection hindering the sample size.









Biological replicate

- 1
- 2
- 3
- 4

Lines

- Wild-Type controls
- Empty vector controls
- Transgenic lines

Figure 4.6: *CMADS1*-RNAi transgenic lines do not show consistent phenotypic differences in frond length, frond branching, or pinnae number compared to controls. The top two plots and bottom left plot each show a different phenotypic measurement of *CMADS1*-RNAi transgenic plants compared with two empty vector transformed controls and an untransformed wild-type control. The bottom left panel offers a comparison using genotyping data of the same lines obtained using qRT-PCR. Normalised *CMADS1* expression was calculated relative to the control wild-type line. Error bars represent standard error. Coloured spots represent biological replicates.

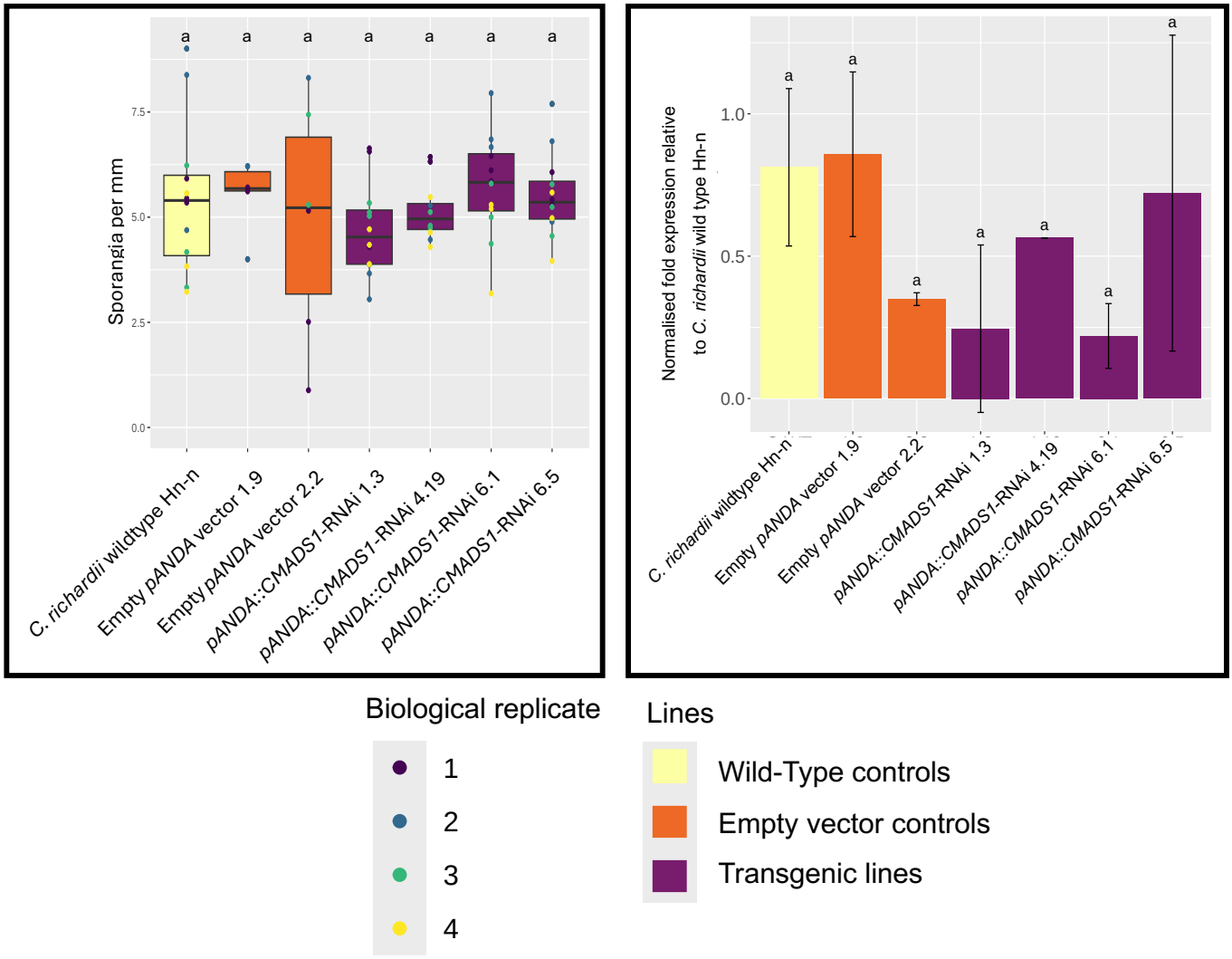


Figure 4.7: *CMADS1*-RNAi transgenic lines do not show consistent phenotypic differences in number of sporangia per mm of pinnae when compared with controls. The left plot shows the phenotypic data of the number of sporangia per mm present in three pinnae from each plant in the line. The right plot shows the genotyping data of the respective lines. Normalised fold expression calculated against control Hn-n wild-type line. Error bars represent standard error. Coloured spots represent biological replicates.

Although the genotyping qPCR of individual lines shows no significant differences between lines, when individual plants within a line are compared significant differences are seen between wild type and empty pANDA vector 2.2 (Figure 4.8). This suggests *CMADS1* expression may have been knocked down in the empty vector control. However, this is much more likely due to a range of expression being observed. Two out of three plants in the *CMADS1-RNAi* 6.1 line and one plant of the *CMADS1-RNAi* 1.3 line have a significantly reduced expression of *CMADS1* when compared to wild-type, but no significant difference from the empty vector control 2.2. However, once again there is no clear correlation between phenotype and genotype (Figure 4.8). Due to the large range observed in the empty vector controls it could be suggested that any stress related to the transformation process is skewing results; this would also rule out any of the lines being true knockdowns of *CMADS1* with further generations needing to be genotyped before any conclusions can be reached.

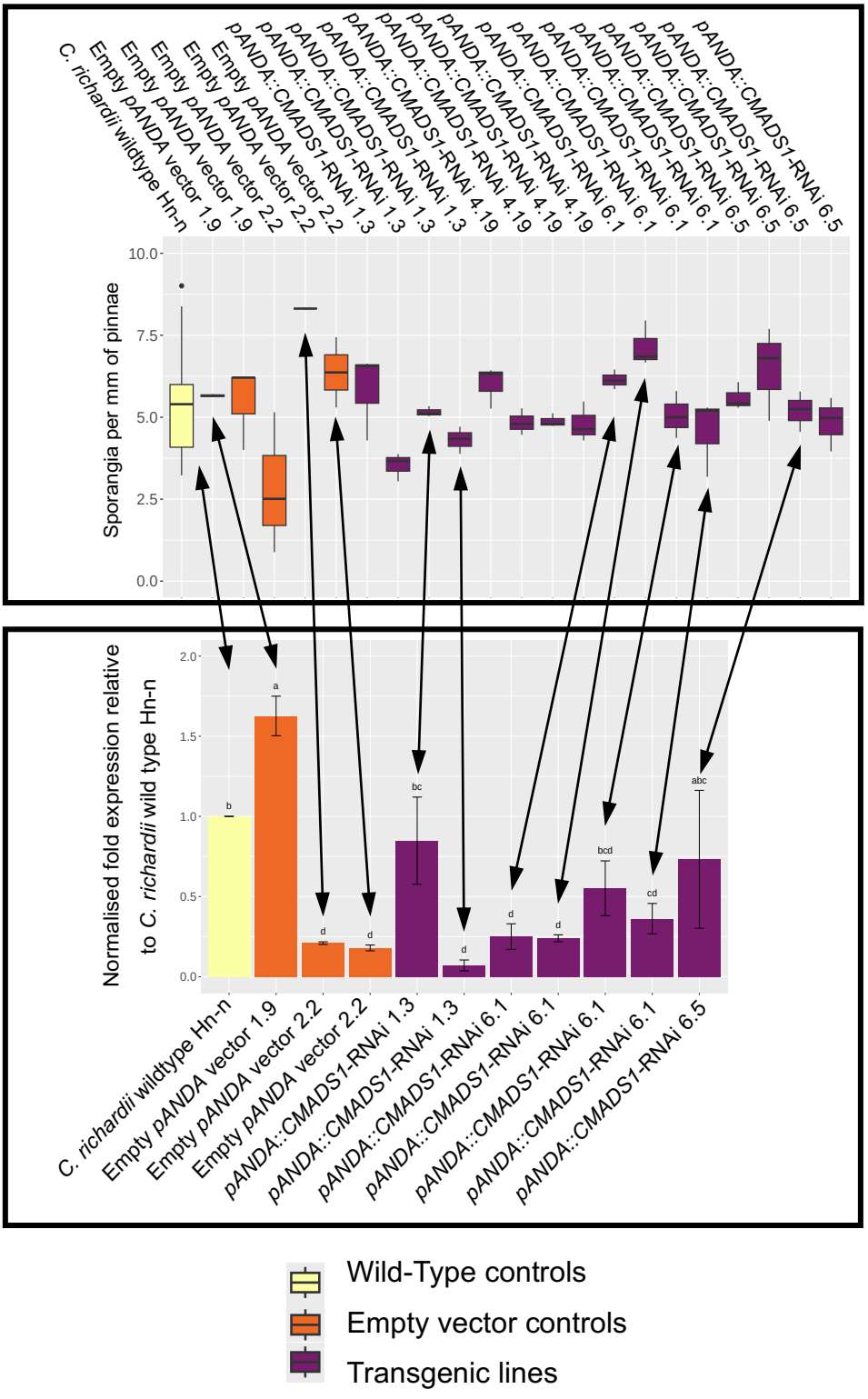


Figure 4.8: Analysis of sporangia per mm and *CMADS1* expression in individual plants. The top plot shows all lines of phenotyped T1 plants. The bottom plot shows a genotyping qPCR representing each individual plant.

4.2.4. Initial characterisation of transgenic *Ceratopteris* plants and selection of tissues for phenotyping

The selected *CMADS1*-RNAi hygromycin-resistant transformants showed no obvious differences in vegetative development compared to wild type controls (Figure 1.9, 4.4).

Figure 4.9: *Ceratopteris* transformants resemble wild type at the vegetative stage of T1 (see Figure 1.9). Images show transgenic vegetative sporophytes photographed 70-days after fertilisation. White bars represent 1 cm.

As it has been suggested that *CMADS1* could have a reproductive function in *Ceratopteris* (Hasebe *et al*., 1997), reproductive phenotypes were assayed in *CMADS1-RNAi* transformants, pANDA empty vector transformants and wild type controls (Figures 4.10-4.15). When describing reproductive sporophytes, morphological changes have occurred compared to vegetative stages, with fronds

curling around sporangia forming pinnae. Fronds branch out, forming branches, these branches then have pinnae, aka, curled fronds. Fiddle heads remain consistent between vegetative and reproductive stages (these can be seen in Figure 4.10 panel B) these are the early stages of frond development and resemble a curled up frond.

To assay reproductive phenotypes, a reproductive frond was selected from each plant. Only one frond was chosen due to the need for spore harvesting at a later time point for continuous spore stocks of the transformation lines. Fronds were chosen when browning was observed of sporangia was observed whilst pinnae still remained green (Figure 4.10 panel D). Frond length was measured, to assay general growth/development, as ancestral MADS-box functions have been suggested to relate to development (Thangavel and Nayar, 2018). Using the harvested frond, counts of total branches per frond and total pinnae per frond were taken as these affect the number of sporangia present in a mature plant. Three pinnae were selected, using the same selecting procedure as described for frond selection. Sporangia were then counted per pinnae to compare potential reproductive viability of *CMADS1-RNAi* plants with wild type *Ceratopteris*.


A
B
C
D
E

Figure 4.10: Hn-n wild type *Ceratopteris* at 138-days in their reproductive sporophyte stage. Phenotyping was performed with measurements taken of frond length, number of branches, number of pinnae, and sporangia per mm of pinnae to act as controls against transformed knockdown *CMADS1*-RNAi lines. Panel A shows the whole plant at the time of harvesting; panel B shows reproductive fiddle heads harvested for genotyping; panel C shows one frond selected for phenotyping; panel D shows three pinnae selected for phenotyping; panel E shows each pinnae dissected to reveal sporangia within. White bars represent 1 cm to scale whilst black represent 1 mm to scale.

Figure 4.11: Line 1.9 empty pANDA vector controls at 138-days old. A frond from each plant was harvested with 3 pinnae then taken for phenotyping under a dissecting microscope. Panel A shows whole plants at time of harvesting; panel B shows fronds selected for phenotyping; panel C shows three pinnae selected for length measurements and sporangia counts; panel D shows dissected pinnae showing sporangia that were then counted. White bars represent 1 cm to scale whilst black represent 1 mm to scale.

Figure 4.12: Line 2.2 empty pANDA vector controls at 138-days old. Panel A shows the fronds selected for phenotyping and panel B shows pinnae dissected for sporangia measurements. White bars represent 1 cm to scale whilst black represent 1 mm to scale.

Whole plant photography was not possible for all lines without damaging plants due to space constraints.

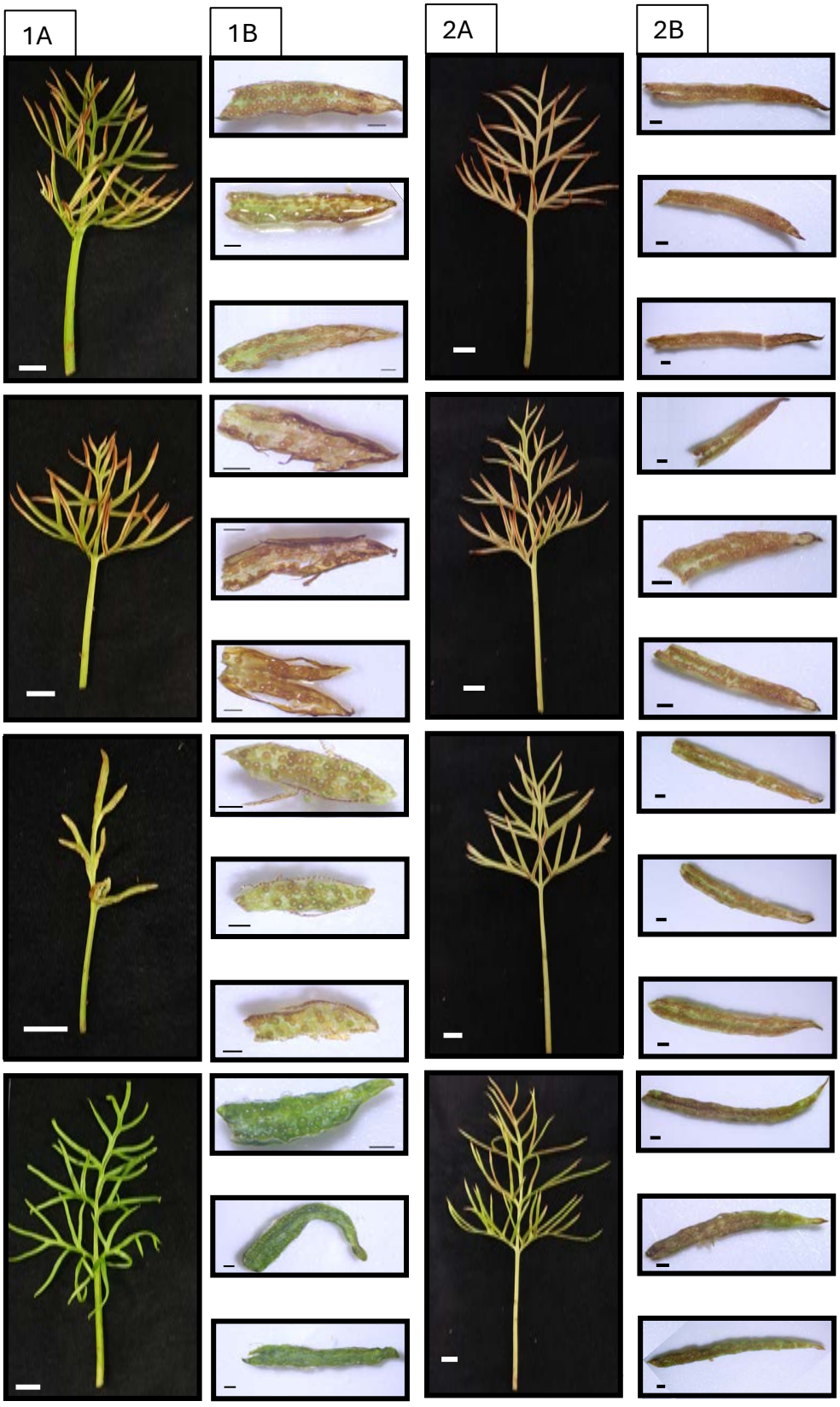

1A
1B
2A
2B

Figure 4.13: Line 1.3 *CMADS1*-RNAi at 138-days (panel 1) and line 4.19 *CMADS1*-RNAi at 138-days (panel 2). Panels A show fronds used for frond length measurements, branch counting, and pinnae counting. Panels B show pinnae dissected to reveal sporangia. White bars represent 1 cm to scale whilst black represent 1 mm to scale.


A
B
C
D

Figure 4.14: Line 6.1 *CMADS1*-RNAi at 138-day. Panel A shows whole plants at 138-days, panel B shows fronds selected for frond length, number of branches, and number of pinnae phenotyping, panel C shows pinnae chosen for pinnae length phenotyping measurements, panel D shows pinnae dissected for sporangia measurements. White bars represent 1 cm to scale whilst black represent 1 mm to scale.


A
B
C

Figure 4.15: Line 6.5 *CMADS1*-RNAi at 138-days. Panel A shows fronds selected for frond length, number of branches, and number of pinnae phenotyping, panel B shows pinnae chosen for pinnae length phenotyping measurements, panel C shows pinnae dissected for sporangia measurements. White bars represent 1 cm to scale whilst black represent 1 mm to scale.

4.3. Overexpression of *CMADS1* in *Ceratopteris* and *Arabidopsis*.

4.3.1. Generation of *CMADS1* overexpression construct for use in *Ceratopteris*

In order to overexpress *CMADS1* in *Ceratopteris*, a 3-step cloning process was undertaken to generate both native and myc-tagged forms of *CMADS1* under the control of the constitutive CaMV 35S promoter (Kiselev *et al*., 2021) in the fern expression vector pBOMBER (Figures 2.2 and 2.3).

However, restriction cloning was delayed due to unsuccessful attempts. Successful insertion of the *CMADS1* sequence into the intermediate vector *pART7* was achieved, but the insertion of *CMADS1* into the delivery vector pBOMBER was not possible in the time required to allow for successful growth of multiple generations of *Ceratopteris*. Trouble shooting was performed as described in Chapter 2. Diagnostic digests can be viewed in the appendix.

4.3.2. *Ceratopteris* transformation of *35S::CMADS1* and *35S::CMADS1*-myc constructs

While the cloning process was occurring, *Ceratopteris* was in parallel bombarded with an empty pBOMBER control vector as in section 4.3.1 and successfully transformed lines were grown to their T1 generation (Appendix 3).

4.3.3 Over expression of *CMADS1* cDNA and *AG* in *Arabidopsis*: generation of constructs.

In order to generate constructs to overexpress *CMADS1* in *Arabidopsis*, a 3-step cloning process was followed to insert *CMADS1* under the control of a *35S* promoter into the binary vector pART27 (Figure 2.4; Aggarwal *et al*., 2018).

4.3.4 Transformation of *Arabidopsis* with *35S::CMADS1* and *35S::AG*

Arabidopsis plants were transformed with both *35S::CMADS1* and *35S::AG* constructs and allowed to set seed. To select transgenic T1 plants, seeds were plated on kanamycin (Figures 4.16 and 4.17). 3 kanamycin-resistant *35S::CMADS1* primary transformant (T1) seedlings and a single *35S::AG* kanamycin-resistant seedling were obtained (Figures 4.16 and 4.17). These 4 plants were propagated onto soil to generate T2 seed and for preliminary phenotypic analysis.

In the T1 generation, two kanamycin-resistant plants were obtained from *35S::AG* line 9. Both plants showed normal vegetative development and flowers that produced all four floral organs, namely sepals, petals, stamens and carpels (Figure 4.18). In the T1 generation of *35S::CMADS1,* line 4 had one kanamycin-resistant plant produced, line 7 had two kanamycin-resistant plants produced, and line 9 had five kanamycin-resistant plants produced (Figure 4.19).

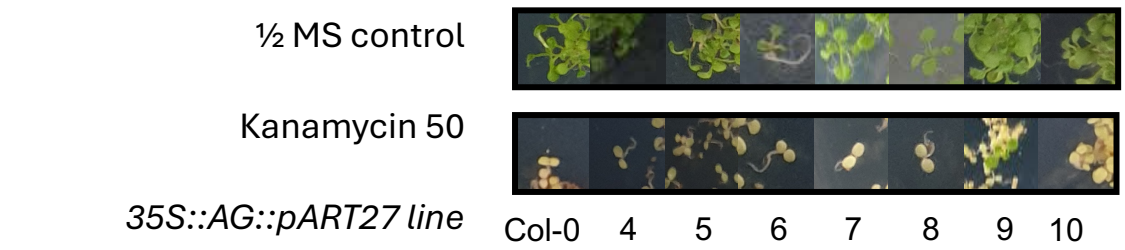


Figure 4.16: T1 lines of *Arabidopsis* transformed with an overexpression *AGAMOUS* insertion undergoing Kanamycin selection. A total of 12 lines were tested for selection with 5 plates being lost to infection. Only a single plant, T1 number 9, was carried forward to soil for propagation and phenotyping.

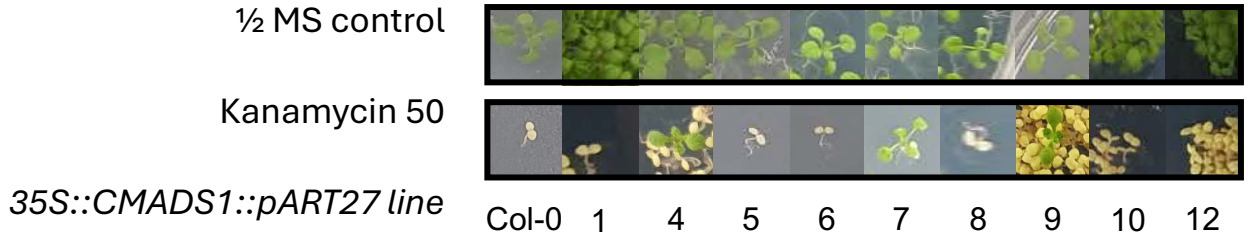
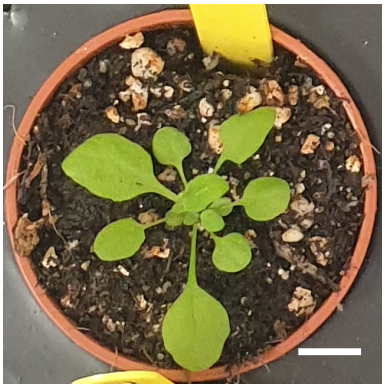


Figure 4.17: T1 lines of *Arabidopsis* transformed with an overexpression *CMADS1* insertion undergoing Kanamycin selection. A total of 12 lines were tested for selection with 3 plates being lost to infection. T1 plants 4, 7 and 9 were kanamycin resistant and were taken forward onto soil for future phenotyping.




35S::AG line 9 plant 1


35S::AG line 9 plant 2

Figure 4.18: Line 9 of T1 *35S::AG Arabidopsis* had two healthy sporophytes successfully grown on Kanamycin selection. These were transferred to soil and had flowers collected for phenotyping and genotyping. Scale bars of sporophytes on soil represent 1 cm to scale, scale bars on flower images represent 1 mm to scale

Heterologous overexpression was performed because AGAMOUS mutants are infertile.


35S::CMADS1 line 4 plant 2


35S::CMADS1 line 7 plant 1


35S::CMADS1 line 7 plant 2


35S::CMADS1 line 9 plant 2


35S::CMADS1 line 9 plant 3


35S::CMADS1 line 9 plant 4


35S::CMADS1 line 9 plant 5

Figure 4.19: **Lines 4, 7, and 9 of T1 35S::CMADS1 Arabidopsis had 1, 2, and 4 respectively, healthy sporophytes successfully grown on Kanamycin selection.** These were transferred to soil and had flowers collected for phenotyping and genotyping. Scale bars of sporophytes on soil represent 1 cm to scale, scale bars on flower images represent 1 mm to scale.

4.4. Discussion

4.4.1. Summary

In this chapter the possibility of cloning the *Ceratopteris* MADS-box gene *CMADS1* for use in both knockdown and overexpression transgenic lines was explored. Exogeneous overexpression of the *Ceratopteris* MADS-box gene *CMADS1* in the model angiosperm *Arabidopsis* was attempted alongside overexpression of the *Arabidopsis* native MADS-box genes *AG.* Long standing transformation techniques have been established for *Arabidopsis* often using *Agrobacterium* (Clough and Bent, 1998; Zhang *et al*., 2020) (see chapter 2). However, the creation of transgenic *Ceratopteris* lines is still undergoing optimisation with microparticle bombardment, Agrobacterium-mediated transformation, and CRISPR/Cas9 all being attempted (Bui *et al*., 2015; Plackett *et al*., 2018; Xiang and Li, 2024).

4.4.2. Transgenic lines of *Ceratopteris* were not reliably produced for either *CMADS1* knockdown or overexpression

Constructs for *CMADS1* knockdown in *Ceratopteris* were successfully generated using Gateway cloning. Antibiotic selection of T1 gametophytes was successful and T2 spores of potential *CMADS1 Ceratopteris* knockdown lines have successfully been collected for future growth to a homozygous generation for phenotyping and

genotyping. Intermediate vectors were also generated that could be transferred to alternative delivery vectors in the future.

In the T1 generation antibiotic selection revealed 3 resistant empty pANDA control lines and 5 resistant *CMADS1-RNAi* lines. Plants that showed antibiotic resistance at the gametophyte stage consistently also showed resistance at the sporophyte stage suggesting stable transformants had been generated. Despite this, expression of *CMADS1* was not consistently significantly knocked down in any of the lines; this could be due to only a partial construct being inserted or to the silenced gene being re-activated, or because passage through another generation is required for a full RNAi effect including via epigenetic promoter modification (Martienssen and Moazed, 2015).

There were no obvious differences in vegetative phenotypes between any of the transformants, empty vector controls or transgenic plants. There were no obvious phenotype differences in the reproductive stage either (Figures 4.10-4.15). The only potentially transgenic line to show a significant difference in phenotype from the wild-type is pANDA::*CMADS1-RNAi* line 6.1 in the number of frond branches. No good correlation between these differences and the *CMADS1* level could be determined. However, the empty pANDA vector control *CMADS1* levels show an increase and a decrease in fold change in the two lines that were successfully taken to the phenotyping stage (Figure 4.6-4.8).

RNAi has previously been used to produce stable *Ceratopteris* transformations (Plackett *et al*., 2018; Withers *et al*., 2023), however these methods relied on protocols established to target angiosperm genes (Bui *et al*., 2017). Previous studies have reported unstable gene silencing when using RNAi in ferns with visible phenotypes being produced in gametophyte stages but re-activation of the gene after fertilisation in the sporophyte stage (Bui *et al*., 2015). However, the two-step antibiotic selection at both the unfertilised gametophyte stage and fertilised sporophyte stage suggest resistance and therefore transformation is stable. Despite RNAi relying on transient expression when used in ferns, rather than stable integration of a transgene it has proved effective in producing phenotypes in ferns (Rutherford *et al*., 2004).

CRISPR/Cas9 methods have recently been optimised in *Ceratopteris* providing faster detection of transgenic plants and an expanded transgenic toolkit in the fern *Ceratopteris* (Jiang *et al*., 2024). However, off target mutations have been observed when using CRISPR/Cas9 in *Ceratopteris* with deletions that could cause frameshifts, premature stop-codons, or removal of diagnostic sites that were only discovered upon sequencing, something that is difficult to do with the large DNA insertions necessary in fern genome editing (Xiang and Li, 2024). Alongside improvements in genetic editing in *Ceratopteris,* an improved method of helium biolistic transformation has been proposed relying on a double barrelled gun approach which has been shown to significantly improve the number of transformants produced in onion (Miller *et al*., 2021).

4.4.3. Overexpression of *CMADS1* in *Ceratopteris* and *Arabidopsis*.

Overexpression in *Ceratopteris* has previously been performed and methods optimised (Youngstrom *et al*., 2022). Constructs for overexpression of *CMADS1* both untagged and myc-tagged were successfully generated, unfortunately due to the timeframe of this project final constructs in the delivery vectors for both *Ceratopteris* and *Arabidopsis* was unsuccessful. In the future these constructs could be used to identify any phenotypes of overexpressed *CMADS1* in both its native *Ceratopteris* and exogenously in *Arabidopsis* where comparisons can be made to overexpression of *AG, CMADS1's* suggested homologue.

Additional future work could focus on developing smaller binary vectors for fern transformation. This has been achieved for *Arabidopsis* through the use of the small binary vector pGREEN with accessory vector pSOUP (Hellens *et al*., 2000; Smidkova, Hola and Angelis, 2010). This would allow for ease of cloning of genes for delivery into fern and potentially other plants.

4.4.4. overexpression lines of *AG* and *CMADS1* were not able to be produced in *Arabidopsis*.

Previous work suggests *CMADS1* could be a homologue of the C-class MADS-box gene AGAMOUS (Hasebe *et al*., 1998). It is difficult to ascertain whether an altered expression of *CMADS1* reflects those observed in *AG* overexpression due to either

unsuccessful transformations or availability of only heterozygous plants. Cloning of *CMADS1* and *AG* into a plant binary vector was achieved as someone else in the lab had managed to clone the constructs described in 4.6.3 into the delivery vector pART27 used for transformation of *Arabidopsis.* However, these constructs have not been checked through sequencing.

An overexpression mutant of *AG* in flowering plants such as *Magnolia* has been shown to give a distinctive phenotype expected of overexpression of a C-class MADS-box gene, namely early flowering and over-development of stamens and carpels. In *Arabidopsis* it has been observed that petals are converted to stamenoid structures (Wu *et al*., 2011). The heterologous overexpression of *Arabidopsis AG* in *Magnolia and* overexpression of *AG* in *Arabidopsis* suggests a conserved phenotype of increased development of stamens and carpels, as *AG* is a C-function gene. One hypothesis is that *CMADS1* could behave in the same way in both *Arabidopsis* and fern, promoting development of female and male reproductive organs.

CHAPTER 5: Do CMADS1 protein binding partners affect CMADS1 function in different tissues?

5.1. Introduction

5.1.1. Overview

The interactions between MADS-box proteins have been well categorised in flowering plants (Puranik *et al*., 2014). Quaternary interactions are responsible for floral organ specification and the whorl structure observed across angiosperms via the floral quartet model (Theissen, 2001) for specifying sepals, petals, stamens and carpels and the later-emerging ABCDE model that encompasses ovule specification (Theissen *et al*., 2016). *AG* has been identified in homodimers and heterodimers when in the formation of a floral quartet tetramer, however the formation of an *AG* homo-tetramer has not been identified (Hsu *et al*., 2013; Theißen, Melzer and Rümpler, 2016 (Figure 1.8). The only MADS-box proteins seen to form homo-tetramers are the E-class proteins *SEPALLATA* (Jetha, Theißen and Melzer, 2014). Yeast-II-Hybrid has been used to produce an interaction map of MADS-box proteins in *Arabidopsis* (de Folter *et al*., 2005) with further Yeast-III-Hybrid assays being performed to determined multimeric complexes formed by MADS-box proteins involved in the ABCDE model (Immink *et al*., 2010).

However, the MADS-box protein family is also present in non-flowering plants that rely on completely different reproductive systems and organs. One way of uncovering the functions of these proteins could be to determine protein function through observing their interactions.

5.1.2. Protein interactions in the MADS-box family throughout the land plant lineage

The MIKC^c subcategory is known for its involvement in flowering time, floral organogenesis, and fruit development (Medard and Yanofsky, 2001). Within the ABCDE model of flowering genes that is used to categorise the MADS-box family in A. thaliana it has been determined that it is the interactions of MADS-box proteins that are responsible for the formation of floral organs (Theissen *et al*., 2016). This interaction between the different MADS-box categories has been well researched in flowering plants (Bowman *et al*., 2012). For example, a complex formed of 2 *AG Class C)* proteins and 2 *SEP* (Class E) proteins will determine carpel identity, and a complex formed of 1 *AG* and 1 *SEP* protein binding with 2 *SHATTERPROOF* or *SEEDSTICK* (class D) proteins will determine ovule identity (Theißen, Melzer and Rümpler, 2016 (Figure 1.4). Thus, loss of function mutants in *Arabidopsis* of Class C MADS-box genes such as *AGAMOUS* lead to a flower with only sepals and petals (see introduction chapter). The genes involved in the ABCDE model evolved before the separation of gymnosperms and angiosperms (Shen *et al*., 2019).

5.1.3. Non-angiosperm MIKC^c protein interactions.

In angiosperms, 14 monophyletic clades have been identified in the MIKC^c subcategory of MADS-box proteins: 11 of these have orthologues in gymnosperms (Chen *et al*., 2017), whereas monilophytes form their own clades as can be seen in Figure 5.4 expanding on what was previously reported (Münster *et al*., 1997). This suggests a divergence of function and possibly interaction partners prior to the advent

of seed plants. In gymnosperms, MADS-box genes are known to be responsible for male, female, and bisexual reproductive structures (Feng *et al*., 2024). Gymnosperm MADS-box proteins also hold a function in the formation of fruits, independent in their evolution from the function attributed to flowers (Chen *et al*., 2017). It has been noted that gymnosperm MADS-box proteins interact similarly to angiosperm MADS-box proteins, yet are more prolific in their interaction as they do not require any Class E proteins to form multimeric complexes and instead can form floral quartet-like complexes alone without the aid of class E proteins in both homo- and hetero-tetramers; this suggests that the floral quartet concept predates the origin of angiosperms (Wang *et al*., 2010).

Within the Bryophyte lineage, mosses contain small MADS-box gene families of an "ancestral" type representing the less complex body architecture (Svensson and Engstrom, 2002). The majority of MADS-box genes present in mosses hold homeotic functions similar to those present in animal MADS-box genes. In the representative bryophyte *Physcomitrium patens* (formerly *Physcomitrella*) the MADS-box family is intermediate in size when compared with the ancestral algae and the dominant angiosperm MADS-box family sizes. *Physcomitrium* has 26 MADS-box genes, 18 of which are Type-II and 6 of which are MIKC^{c} (Rensing *et al*., 2007). In *Physcomitrium* it has been shown that some MADS-box functions are conserved between mosses and the flowering plant lineage, such as the regulation of cell division, whilst other functions have been lost, such as those relating to sperm motility, which was lost along with the loss of flagellated sperm as plants diversified to produce flowering organs (Koshimizu *et al*., 2018). It has been hypothesised that due to the presence of a K-box domain, the domain responsible for protein-protein interactions, in every MADS-box

gene identified in *P. patens* (Barker and Ashton, 2013), that the MADS-box proteins present undergo similar protein-protein interactions to those observed in flowering plants forming complex gene regulatory networks (Henschel *et al*., 2002). Liverworts have smaller families of MADS-box genes, for example, *M. polymorpha* has only two identified MADS-box genes (Bowman *et al*., 2017), although homo-dimerisation of *Mp*MADS1 has been observed suggesting even in such small networks, protein interaction is still integral to MADS-box function (Zobell *et al*., 2010).

5.1.4. Aims and objectives of this chapter

The aim of this chapter was to explore potential protein-protein interactions of *Ceratopteris* MIKC^c MADS-box proteins. The objectives of the work were to:

1. Determine similarities between *Arabidopsis* and *Ceratopteris* Type-II MIKC^c MADS-box sequences to explore the extent of conservation between lineages.
2. Observe any differences in the putative phylogenetic relationships of *Arabidopsis* and *Ceratopteris* Type-II MIKC^c MADS-box proteins, based on analysis of the most conserved and least conserved domains.
3. Predict secondary structures of *Ceratopteris* Type-II MIKC^c MADS-box proteins and compare with the known *Arabidopsis* proteins.
4. Analyse interactions between selected *Ceratopteris* MADS-box proteins using the yeast-II-hybrid system.

5.2. Conserved region comparison of protein sequences within the MADS-box family.

5.2.1. *Ceratopteris* MADS-box proteins conserved regions

In order to determine conserved protein sequence regions in *Ceratopteris* MADS box proteins, a multiple sequence alignment was performed on all full-length *Ceratopteris* Type-II MIKC^c proteins (Figure 5.1). Full length sequences were used to ensure any potential conservation could be detected in the I and C-domains as well as in the better-characterised M and K domains.

The alignment is concurrent with previous literature from other species that describes the M-domain as the most conserved domain in the MADS-box family, with the K-domain showing conservation to a lesser degree (Figure 5.1). Between the M and K domains the Intervening (I) domain sits showing potential low-level conservation, the C-terminus C-domain shows the least conservation of the domains (Theißen, Kim and Saedler, 1996) (Figure 5.1).

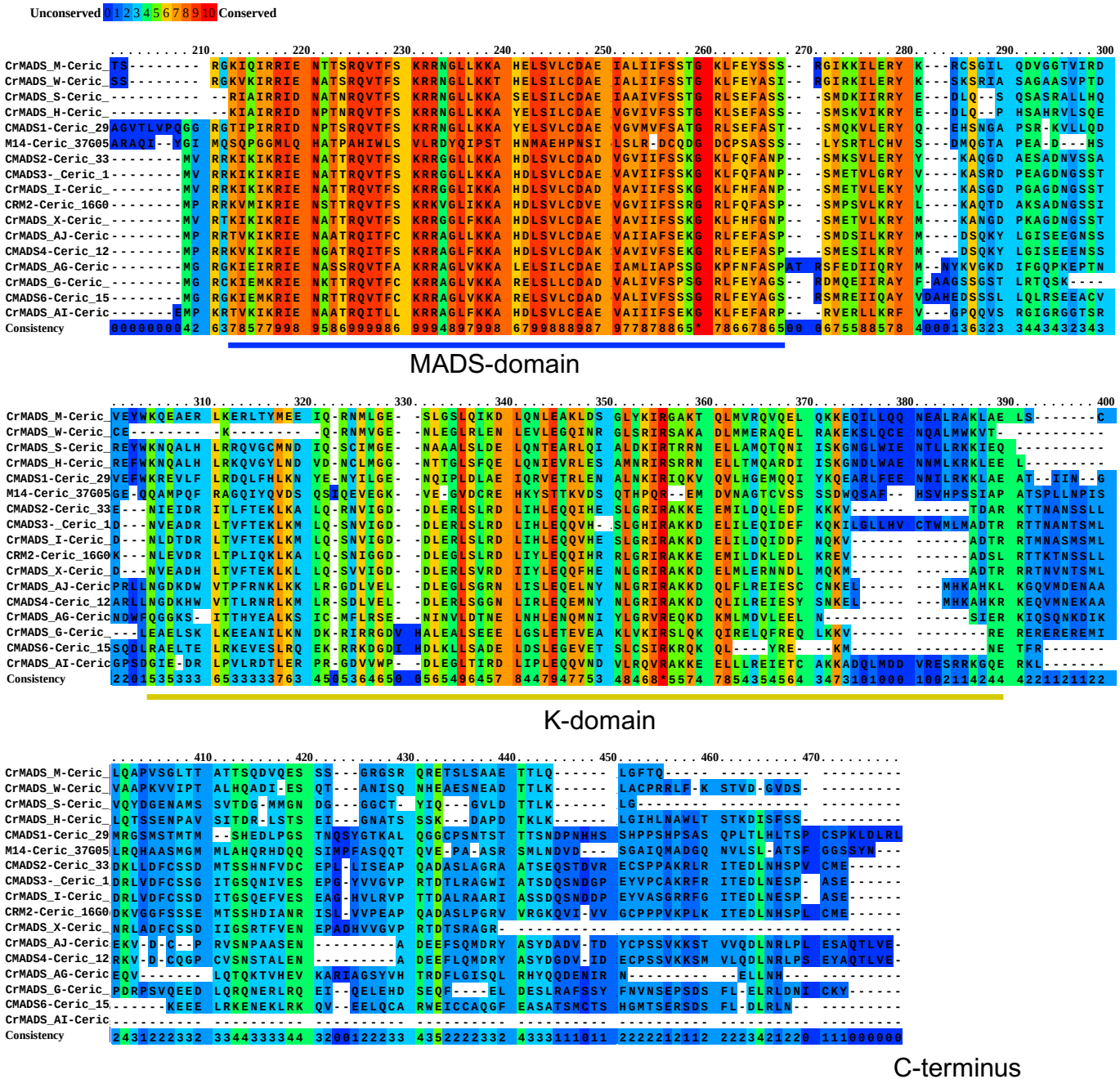


Figure 5.1: multiple sequence alignment (MSA) of all identified *Ceratopteris* Type-II MIKC^c MADS-box proteins using PRALINE MSA (Simossis and Heringa, 2003). Most conserved amino acids are coloured red with least conserved coloured blue. The M domain is underlined in blue and labelled, the K domain is underlined in yellow and labelled, the K1-3 sub-domains can be seen in the conservation patterns. Red highlight represents high conservation, blue represents low conservation, colours are on a scale of 1-10 with 10 showing the highest level of conservation whilst 0 represents the least.

5.2.2. *Arabidopsis* MADS-box protein conserved regions.

To compare conserved protein sequence regions between *Ceratopteris* MADS box proteins and *Arabidopsis* MADS box proteins, an additional multiple sequence alignment was initially performed on all full-length *Arabidopsis* Type II MIKC^c proteins (Figure 5.2).

This alignment reflects the known levels of conservation throughout the MADS-box family in angiosperms with the conservation of domains from highest to lowest being M, K, I, and C domains (Theißen, Kim and Saedler, 1996). This MSA also shows the three alpha helices present in the K-domain (Figure 5.2, 1.6). This structural motif has been attributed to the function of protein-protein binding (Kaufmann, Melzer and Theißen, 2004).

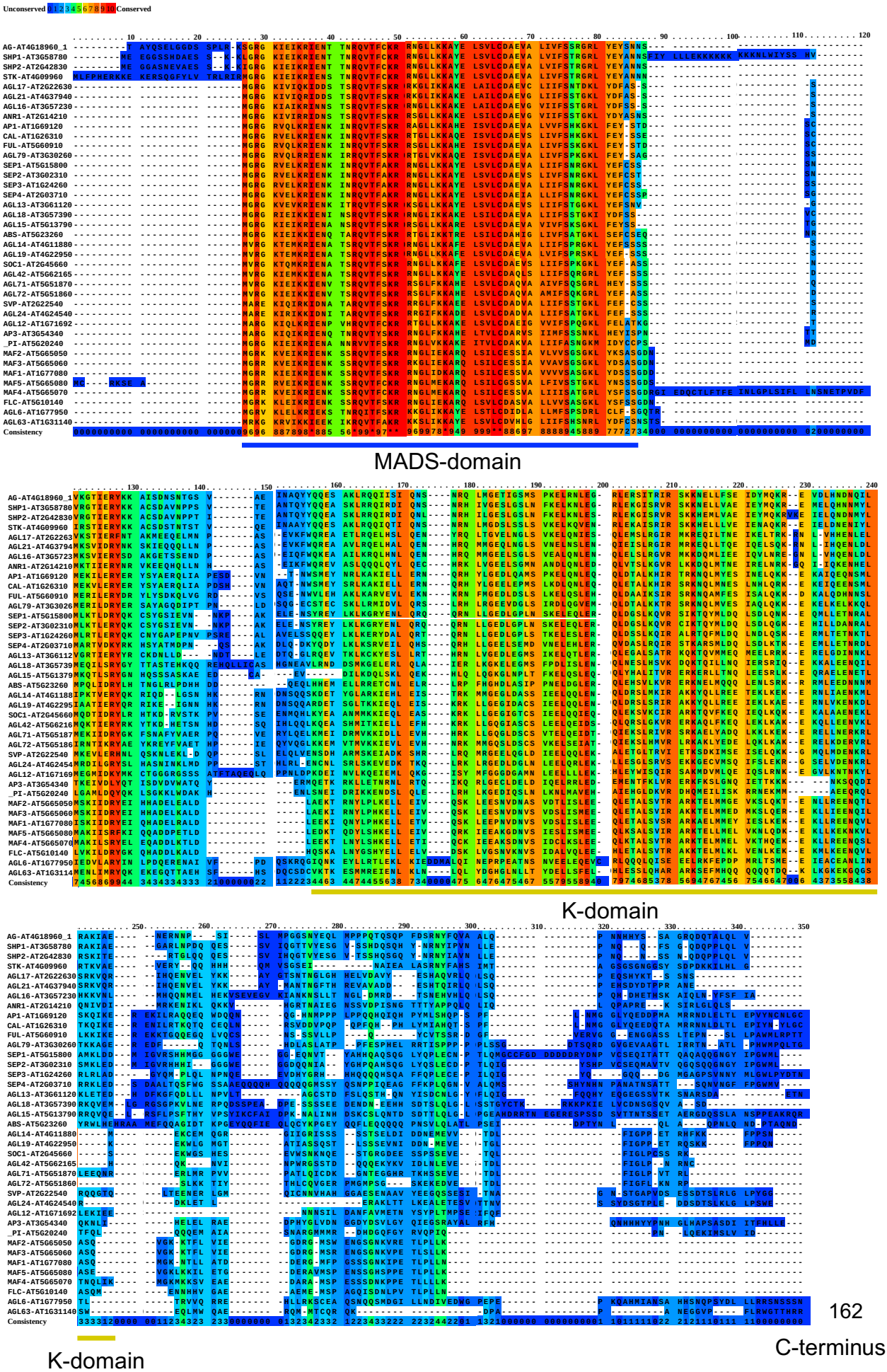
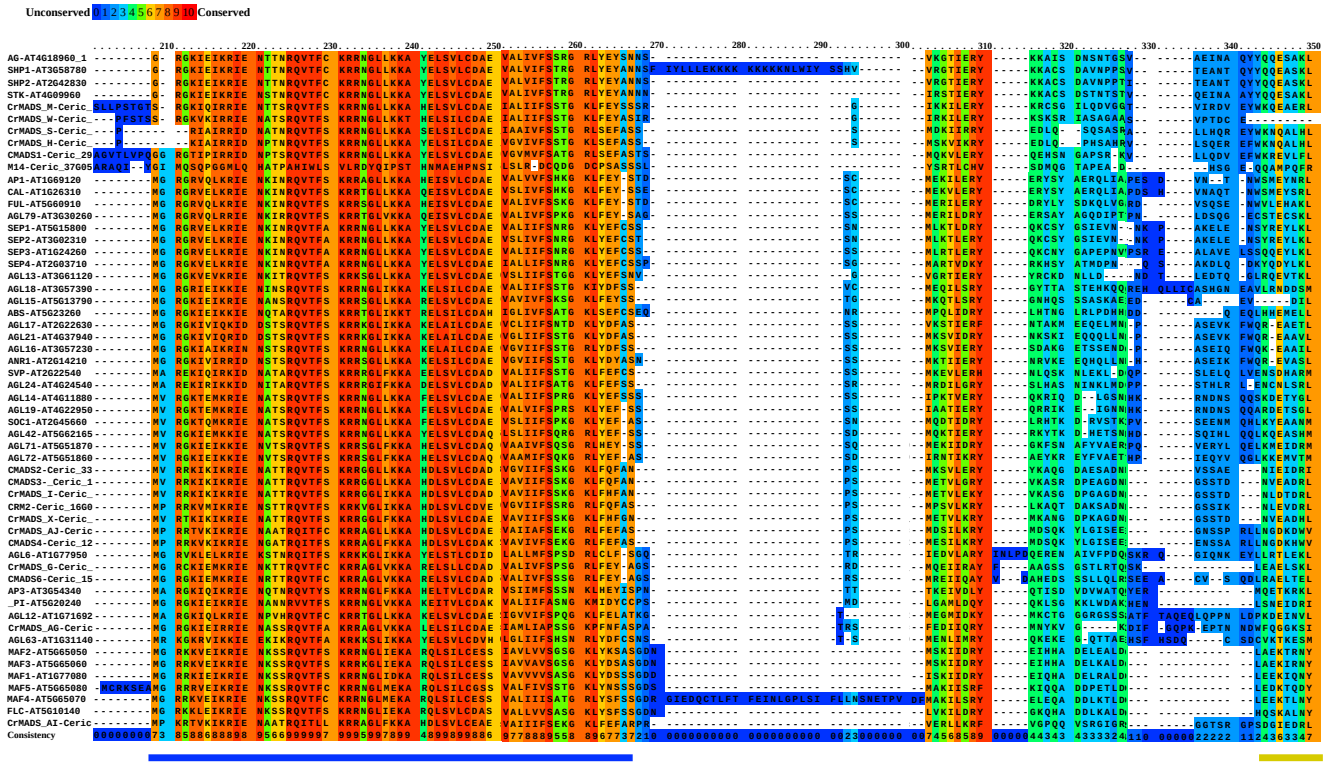

Unconserved 1 2 3 4 5 6 7 8 9 10 Conserved
MADS-domain
K-domain
K-domain
C-terminus
162

Figure 5.2: MSA of all identified *Arabidopsis* Type-II MIKC^c MADS-box proteins using PRALINE MSA (Simossis and Heringa, 2003). Most conserved amino acids are coloured red with least conserved coloured blue. The M domain is underlined in blue and labelled, the K domain is underlined in yellow and labelled, the K1-3 sub-domains can be seen in the conservation patterns. Red highlight represents high conservation, blue represents low conservation, colours are on a scale of 1-10 with 10 showing the highest level of conservation whilst 0 represents the least.

5.2.3. Sequence alignment of MADS-box proteins between *Ceratopteris* and *Arabidopsis*.

To further compare conserved domains of MADS-box proteins from *Ceratopteris* and *Arabidopsis,* a further MSA was performed encompassing both *Ceratopteris* and *Arabidopsis* sequences (Figure 5.3). The alignment shows the same conservation pattern the as species' individual alignments, suggesting the MADS-box domains have been conserved throughout plant evolution. The secondary structure of the K-domain is less noticeable in this alignment than it is in the *Arabidopsis* alone alignment (Figure 5.2) and shows a similar conservation pattern as that shown in the *Ceratopteris* MSA (Figure 5.1), despite there being fewer sequences in *Ceratopteris*. Taken together these alignments show that *Ceratopteris* MADS-box proteins hold similar conserved domains to flowering plant MADS-box proteins. However, the K-domain alignment suggests less conservation between *Ceratopteris* proteins than the conservation observed in the *Arabidopsis* K-domain.


Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved
MADS-domain
K-domain

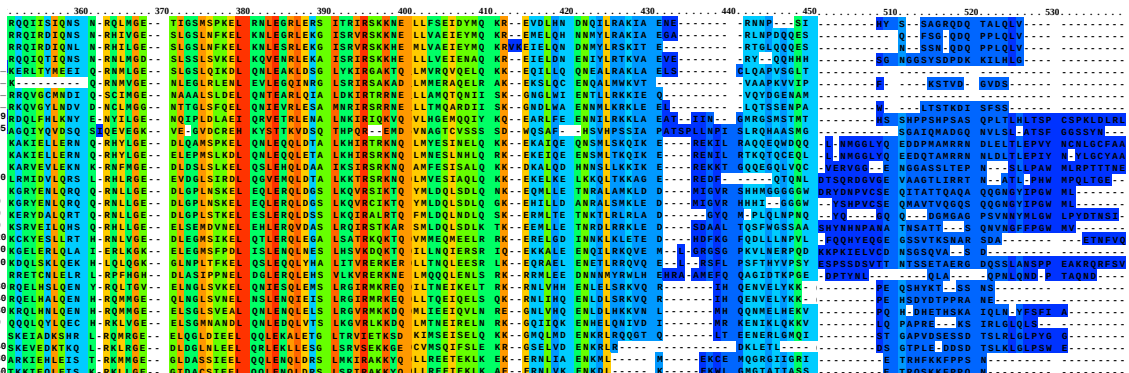

K-domain
C-terminus

Figure 5.3: MSA of all identified *Arabidopsis* and *Ceratopteris* Type-II MIKC^c MADS-box proteins using PRALINE MSA (Simossis and Heringa, 2003). Most conserved amino acids are coloured red with least conserved coloured blue. The M domain is underlined in blue and labelled, the K domain is underlined in yellow and labelled, the K1-3 sub-domains can be seen in the conservation patterns. Red highlight represents high conservation, blue represents low conservation, colours are on a scale of 1-10 with 10 showing the highest level of conservation whilst 0 represents the least.

5.3. Comparison phylogenetic trees of MIKC^c proteins.

5.3.1. Full length protein comparisons.

Following the protein MSAs, phylogenies were produced using full-length proteins to compare relationships between *Ceratopteris* MADS-box Type-II MIKC^c proteins and the clades they form and those formed by *Arabidopsis* MADS-box Type-II MIKC^c proteins (Figure 5.4). A MIKC* protein from *Ceratopteris* and a MIKC* protein from *Arabidopsis* were used as an outgroup for rooting the tree. The known clades of *Arabidopsis* MIKC^c proteins are formed clearly, reflecting those formed in Figure 3.1 when representative species from each major plant group are included. *Ceratopteris* MIKC^c proteins occupy their own clades as previously described (Münster *et al*., 1997) (Figure 3.1). This full-length protein phylogeny suggests the closest *Ceratopteris* relatives to the AG clade are the proteins that fall into the CMADS6 clade namely CMADS6 and CrMADS_G.

Only two *Ceratopteris* MADS-box proteins, *Cr*MADS_M and *Cr*MADS_W (other than the two MIKC* proteins used as an outgroup), fall into an *Arabidopsis* clade sitting in the AGL15 clade. Largely the *Ceratopteris* MADS-box proteins form their own monophyletic clades named in this chapter (and Chapter 3) as: CMADS6, CMADS1, and CMADS3. Previous phylogenies generated in this thesis consistently show CrMADS_G falls within the same clade and is closely related to CMADS6. Previously described CRM1, CRM3, and CRM6 clades have been named on this tree as CMADS3, CMADS6, and CMADS1 respectively, as the nomenclature has progressed.

As more MADS-box genes are identified within *Ceratopteris*, duplicate names for the same genes are inevitable as is seen in *Arabidopsis*.

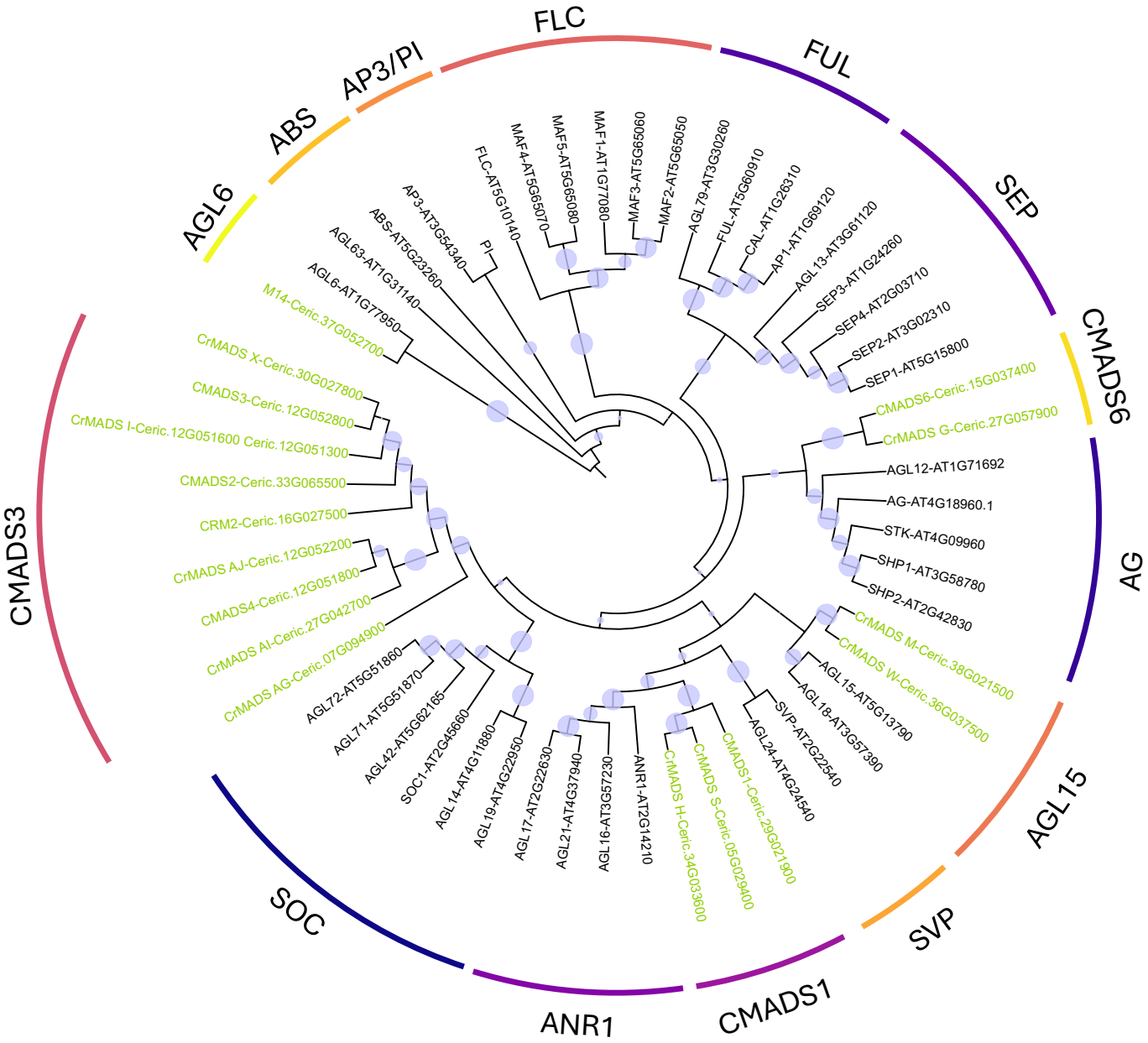


Figure 5.4: Phylogeny of all identified Type-II MIKC^c MADS-box proteins of *Arabidopsis* and *Ceratopteris*. Full-length sequences were aligned using Clustal Omega MSA (Madeira *et al*., 2024). A phylogenetic protein tree was constructed based on the maximum likelihood method (McCormack, Huang and Knowles, 2009). Protein names are represented by the most used name as well as the accession numbers for the sequences. Purple spheres on some branches represent bootstrap values, the larger the sphere the higher the bootstrap, the highest bootstrap was: 1 and the lowest was: 0.48. *Arabidopsis* proteins are presented in black text, *Ceratopteris* proteins are presented in green text. Based on this tree Type-II MIKC^c proteins in *Arabidopsis* and *Ceratopteris* can be categorised into 14 clades.

The full-length protein sequences were used in this tree with all four domains, M, I, K, and C, being present. Previous phylogenies of *Arabidopsis* MADS-box proteins often removed the lesser conserved regions (Gramzow and Theißen, 2015). As such, phylogenetic trees of truncated *Ceratopteris* and *Arabidopsis* proteins were generated.

5.3.2. Truncated MIK protein phylogeny.

The C-terminus was removed from the sequences using the known conserved K-domain sequences from *Arabidopsis* MADS-box proteins along with the alignments presented in this chapter (Figure 5.3) to determine where to truncate the proteins. Similar clades were observed as those seen in the full-length protein tree (Figure 5.4), with the majority of differences between trees being in the deep evolutionary branches that have lower bootstrap values. The actual clades present reflect those seen in the full-length protein phylogeny, suggesting that even with the less conserved amino acids being removed the inferred relationships between the proteins remain the same (Figure 5.5).

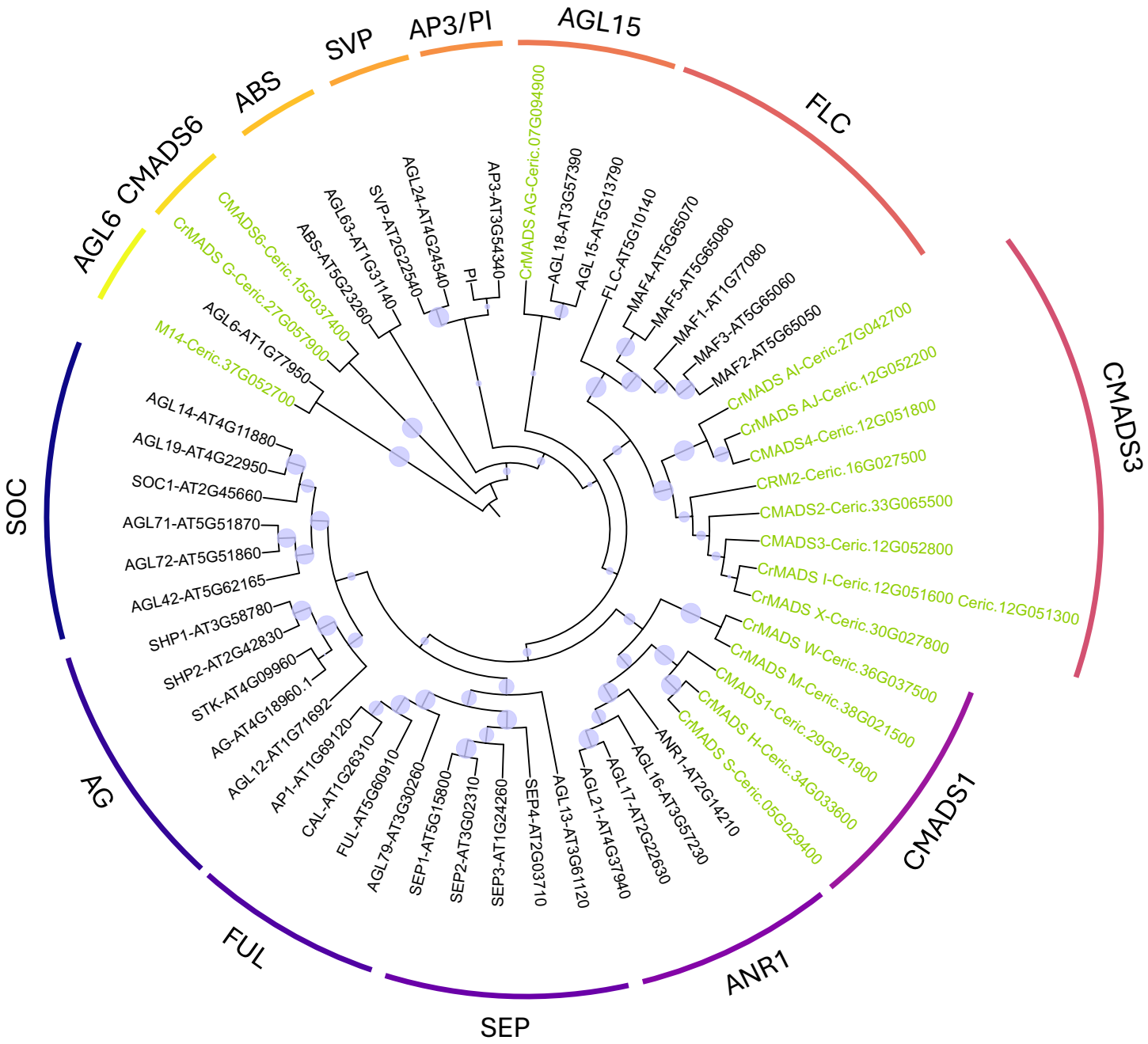


Figure 5.5: Phylogeny of all identified Type-II MIKC^c MADS-box proteins of *Arabidopsis* and *Ceratopteris* using the MIK domains only. Proteins were aligned and the tree produced in the same way described in figure 5.4. The highest bootstrap was 1 and the lowest was: 0.33. *Arabidopsis* proteins are presented in black text, *Ceratopteris* proteins are presented in green text. Based on this tree, Type-II MIKC^c can be categorised into 14 clades.

5.3.3. Truncated M domain protein phylogeny.

A final phylogeny was performed using only the highly conserved M domains defined in Figure 5.3. Based on this tree, Type-II MIKC^c proteins can be categorised into 11 clades due to merging of the AG and CMADS6 clades, SEP and FUL clades, and the FLC AGL15 and PI clades. The CMADS6 and AG clade are neighbouring in the full-length protein phylogeny and merged in the M-domain only phylogeny. This suggests that the closest relation to the AG clade is the CMADS6 clade despite the proteins within representing *Arabidopsis* and *Ceratopteris* respectively.

The lowest bootstrap value (0.17) was reported on the M-domain only phylogeny. This indicates that robust protein sequence comparisons require interrogation of both conserved and less-conserved regions, most likely as all regions contribute to the overall protein structure (Figure 5.6).

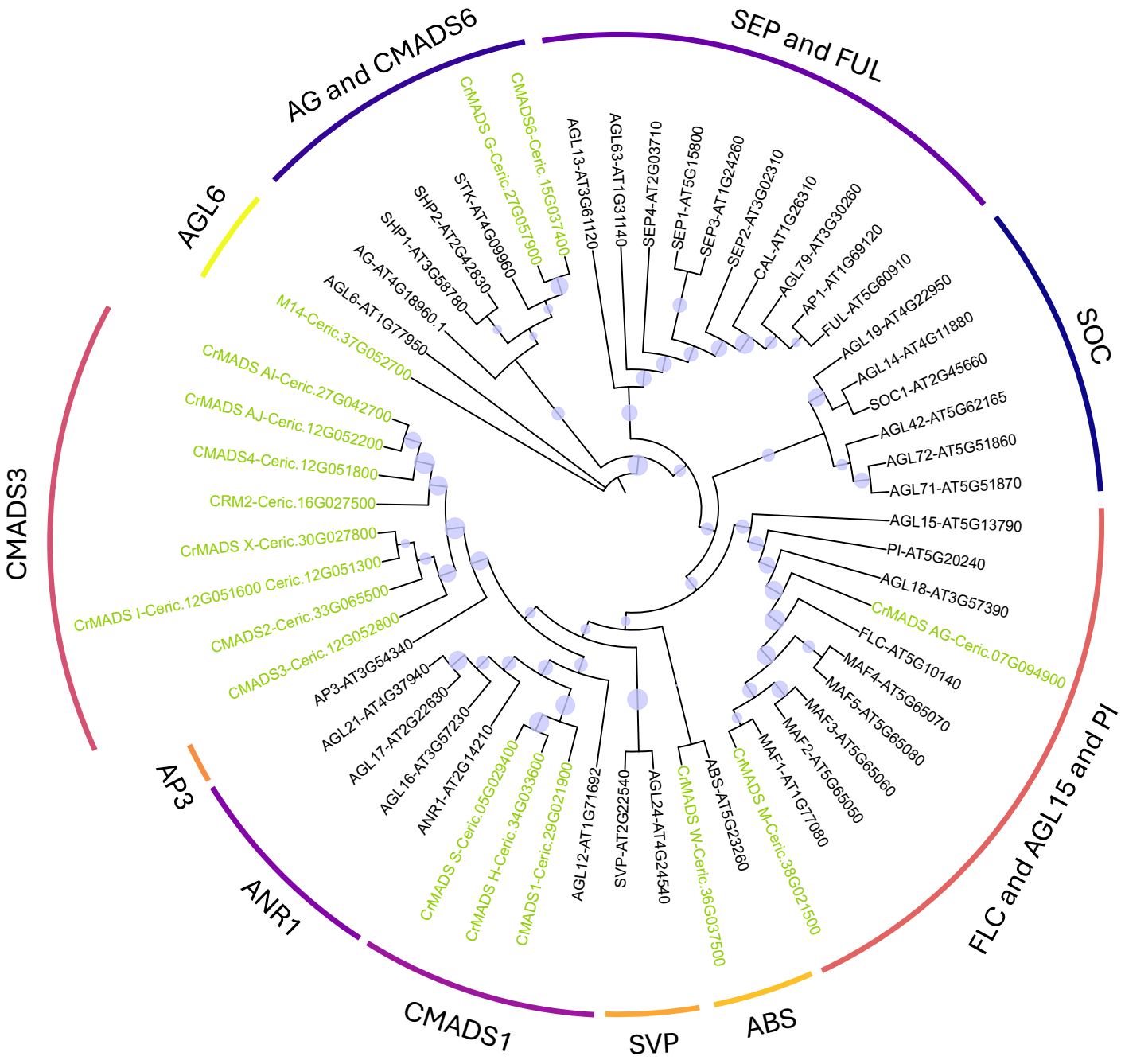
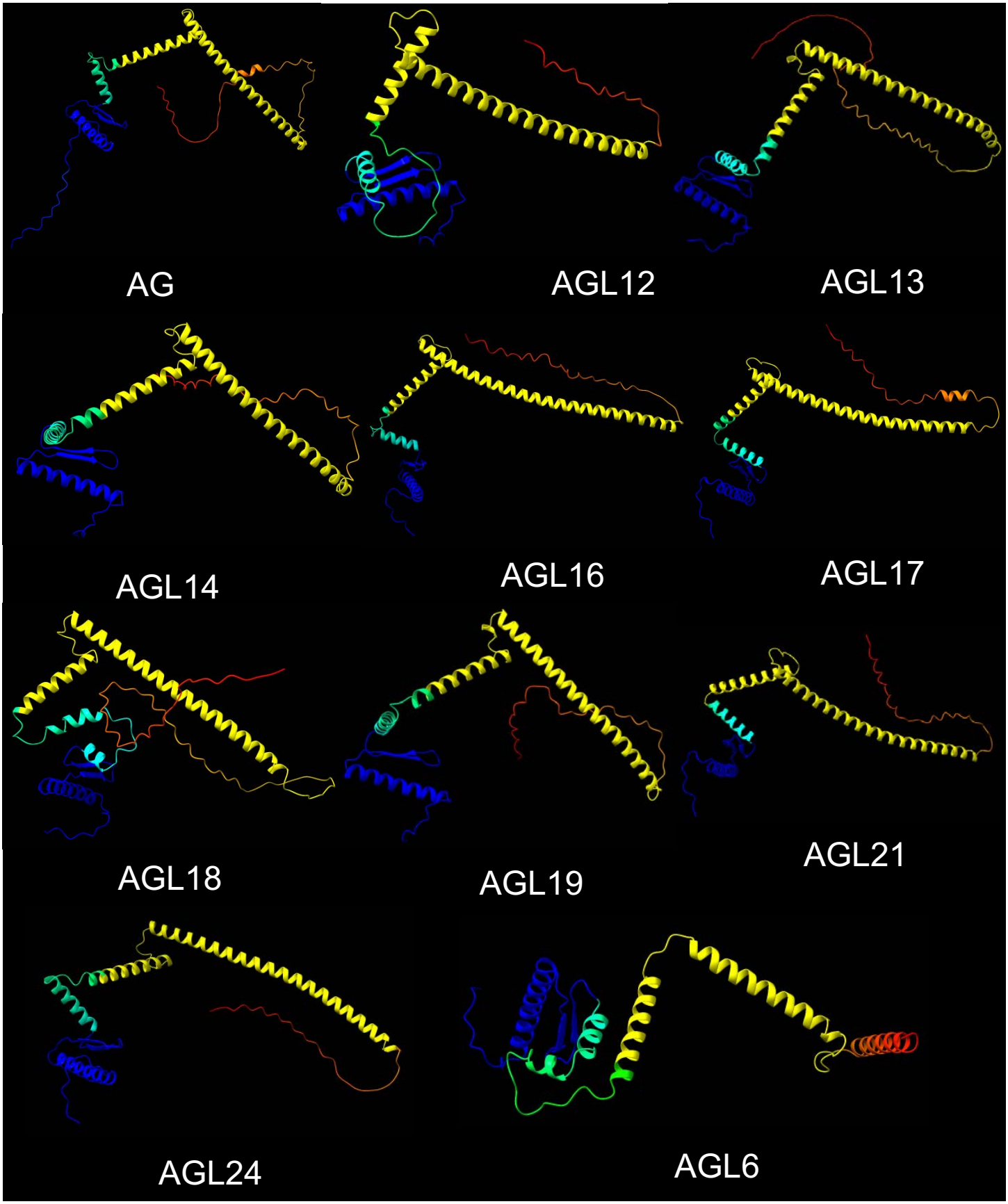


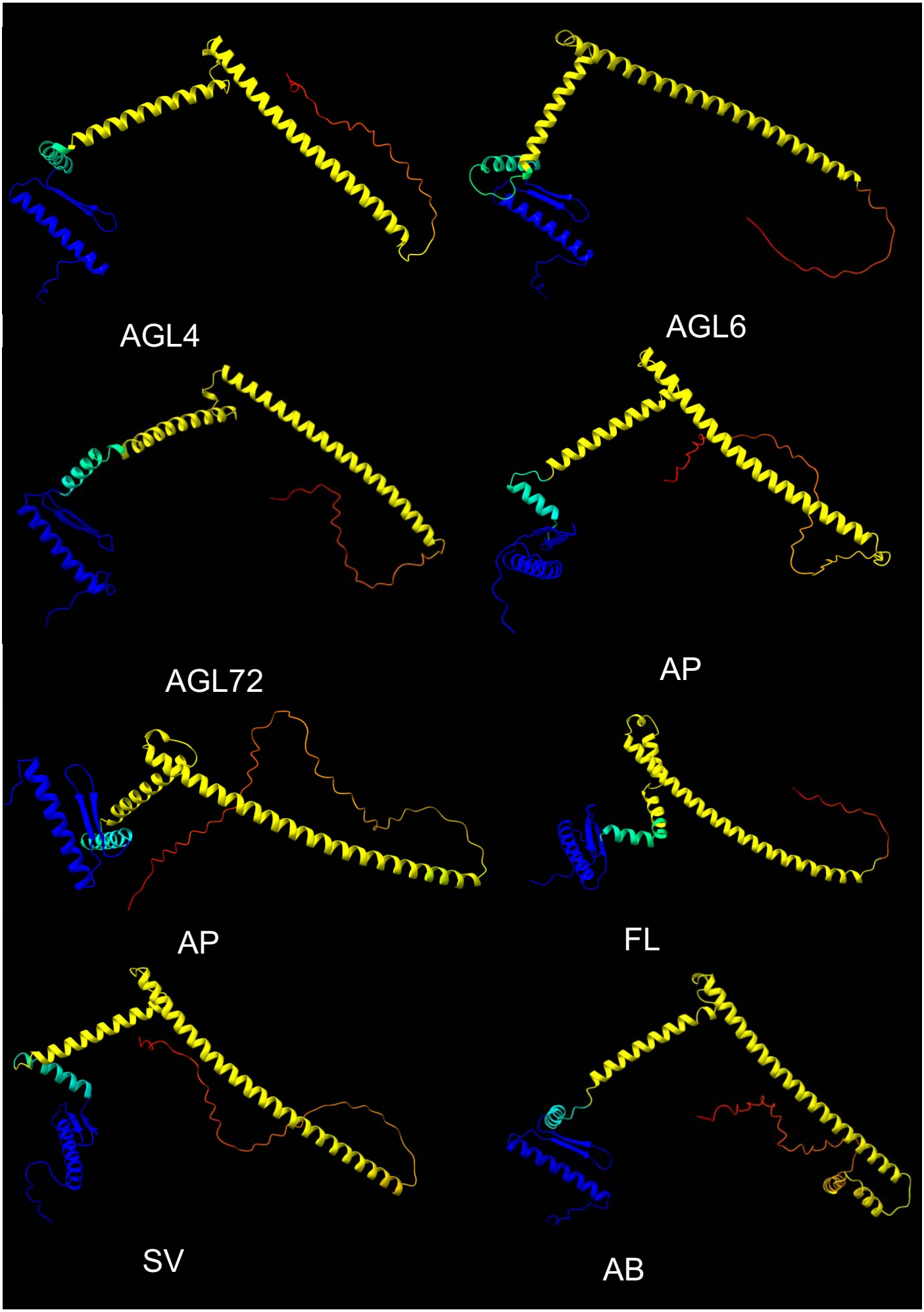
Figure 5.6: Phylogeny of all identified Type-II MIKC^c MADS-box proteins of *Arabidopsis* and *Ceratopteris* using only the M domain. Proteins were aligned and the tree produced in the same way described in figure 5.4. The highest bootstrap was: 0.96 and the lowest was: 0.17. *Arabidopsis* proteins are presented in black text, *Ceratopteris* proteins are presented in green text.

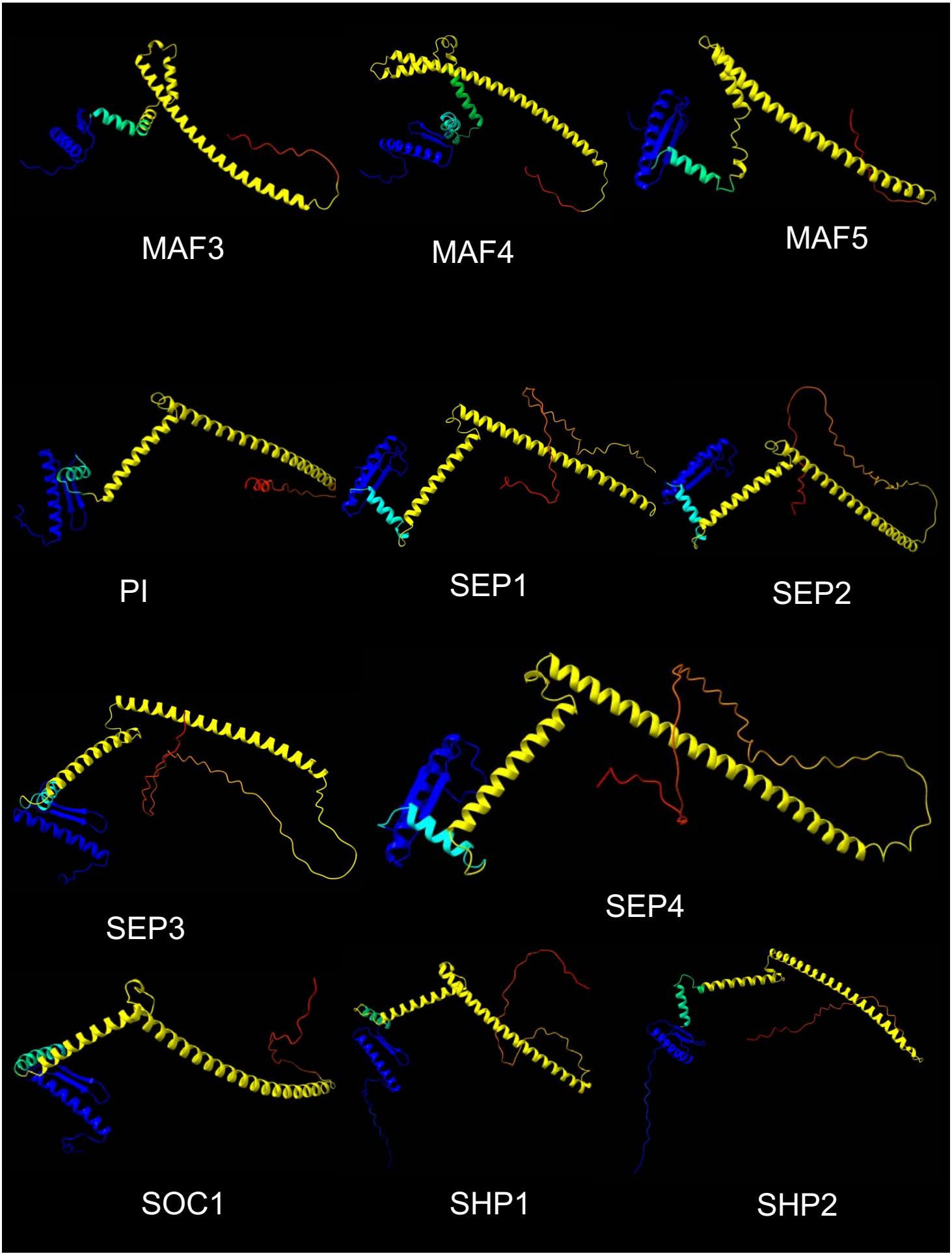
5.4. Protein modelling of MADS-box proteins in *Ceratopteris* and *Arabidopsis*.

5.4.1. Protein secondary structure predictions of *Arabidopsis* MADS-box proteins.

Secondary structures of *Arabidopsis* MADS-box proteins have largely been predicted with the crystal structure of SEP3 resolved (Qui *et al*., 2023). To further investigate *Arabidopsis* MADS-box protein structures, all Type-II MIKC^c protein sequences from *Arabidopsis* were modelled to highlight the predicted structure of their conserved M and K domains (Figure 5.7). There is conservation of structural motifs such as the short alpha helices and two beta sheets that make up the M-domain and the two longer alpha helices that make up the K-domain (Figure 5.7). There is high predicted structural similarity between all *Arabidopsis* MADS-box Type-II MIKC^c *Arabidopsis* proteins, suggesting a conserved structure relevant to their function (Figure 5.7).


AG
AGL12
AGL13
AGL14
AGL16
AGL17
AGL18
AGL19
AGL21
AGL24
AGL6


AGL4
AGL6
AGL72
AP
AP
FL
SV
AB


MAF3
MAF4
MAF5
PI
SEP1
SEP2
SEP3
SEP4
SOC1
SHP1
SHP2

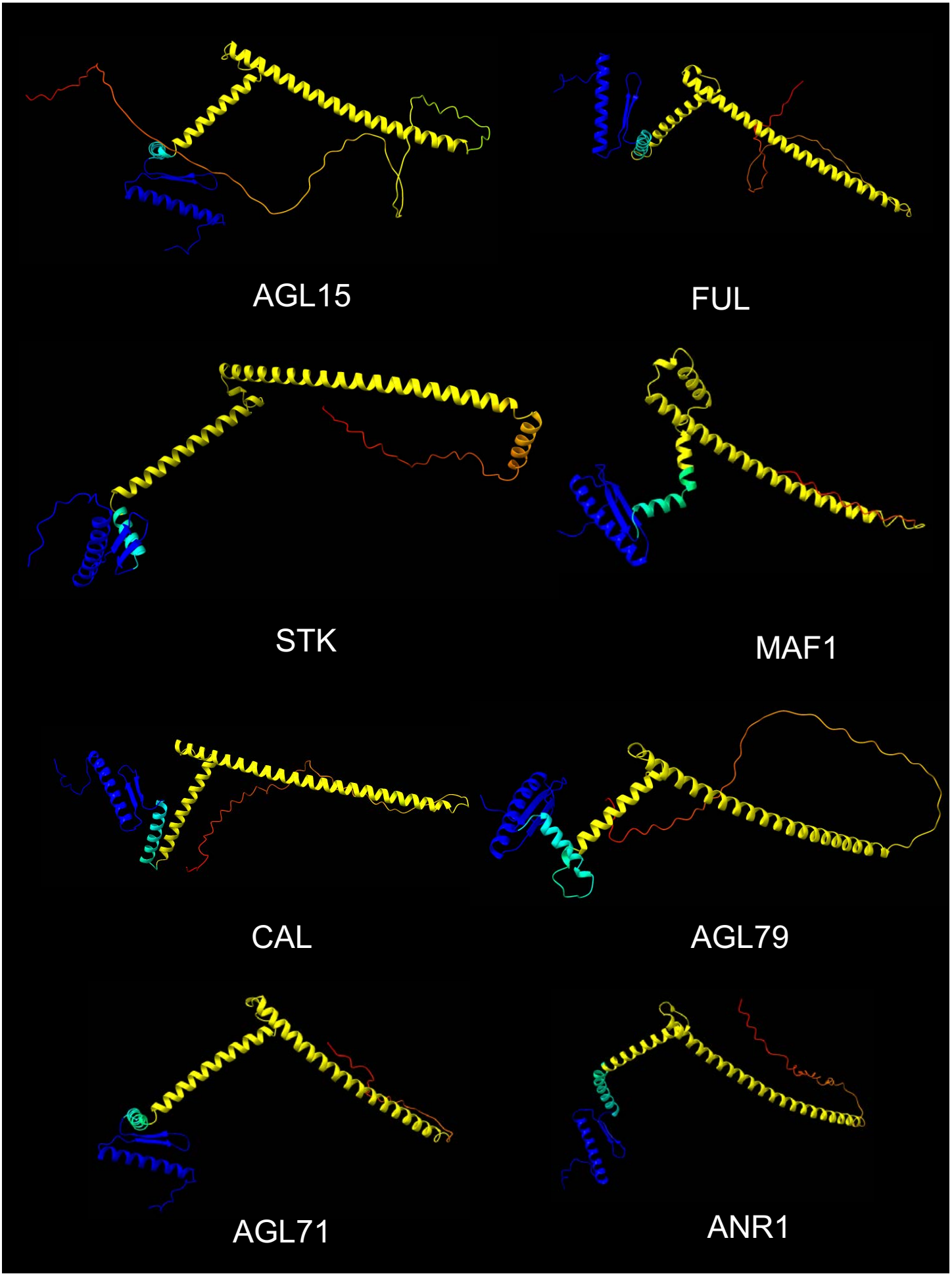
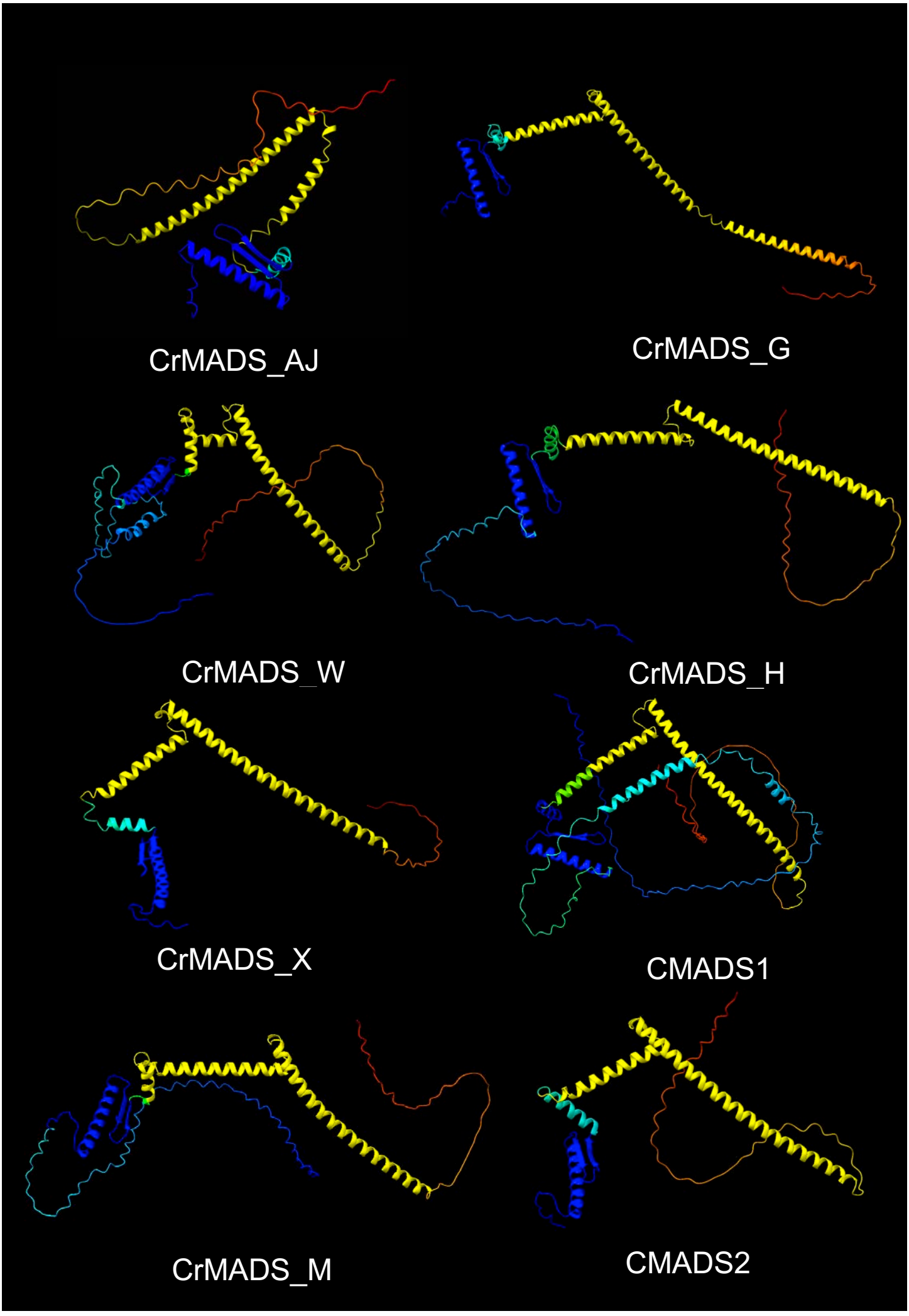

AGL15
FUL
STK
MAF1
CAL
AGL79
AGL71
ANR1

Figure 5.7: structural modelling of *Arabidopsis* MADS-box proteins. All identified Type-II MIKC^c MADS-box proteins in *Arabidopsis* were visualised using ChimeraX (UCSF ChimeraX, 2023). The blue sections represent the short alpha helix and two beta sheets of the M domain, the yellow sections represent the two alpha helices (one longer than the other) of the K domain.

5.4.2. Modelling of *Ceratopteris* Type-II MIKC^c MADs-box proteins secondary structures.

In order to predict the secondary structures of *Ceratopteris* MADS-box proteins and compare them to their Arabidopsis counterparts, MIKC^c protein sequences from *Ceratopteris* were modelled using the same methods as the *Arabidopsis* proteins (section 5.4.1).

There is more divergence between *Ceratopteris* MADS-box protein secondary predicted structures than amongst the *Arabidopsis* proteins (Figures 5.7 and 5.8). In particular, *Cr*MADS_H and CMADS6 have differing structures for their K-domain alpha helices predicted tertiary structures This could be due to overall sequence divergence leading to the model not being the best fit. This could also suggest the secondary structures are less conserved than those observed in *Arabidopsis* due to potential divergent functions. During the course of this work it was discovered that *Cr*MADS_G and *Cr*MADS_AI are both part of the same gene, with G encoding the M domain and AI encoding the K domain, despite being annotated in the genome as 2 separate genes (Marchant *et al*., 2019). This has been taken into account in Figure X below. The secondary structures of *Cr*MADS_AJ and *Cr*MADS_L were not able to be resolved using this method, however, they do still contain the conserved motifs of the M-domain beta sheets and the K-domain alpha helices (5.8).


CrMADS_AJ
CrMADS_G
CrMADS_W
CrMADS_H
CrMADS_X
CMADS1
CrMADS_M
CMADS2

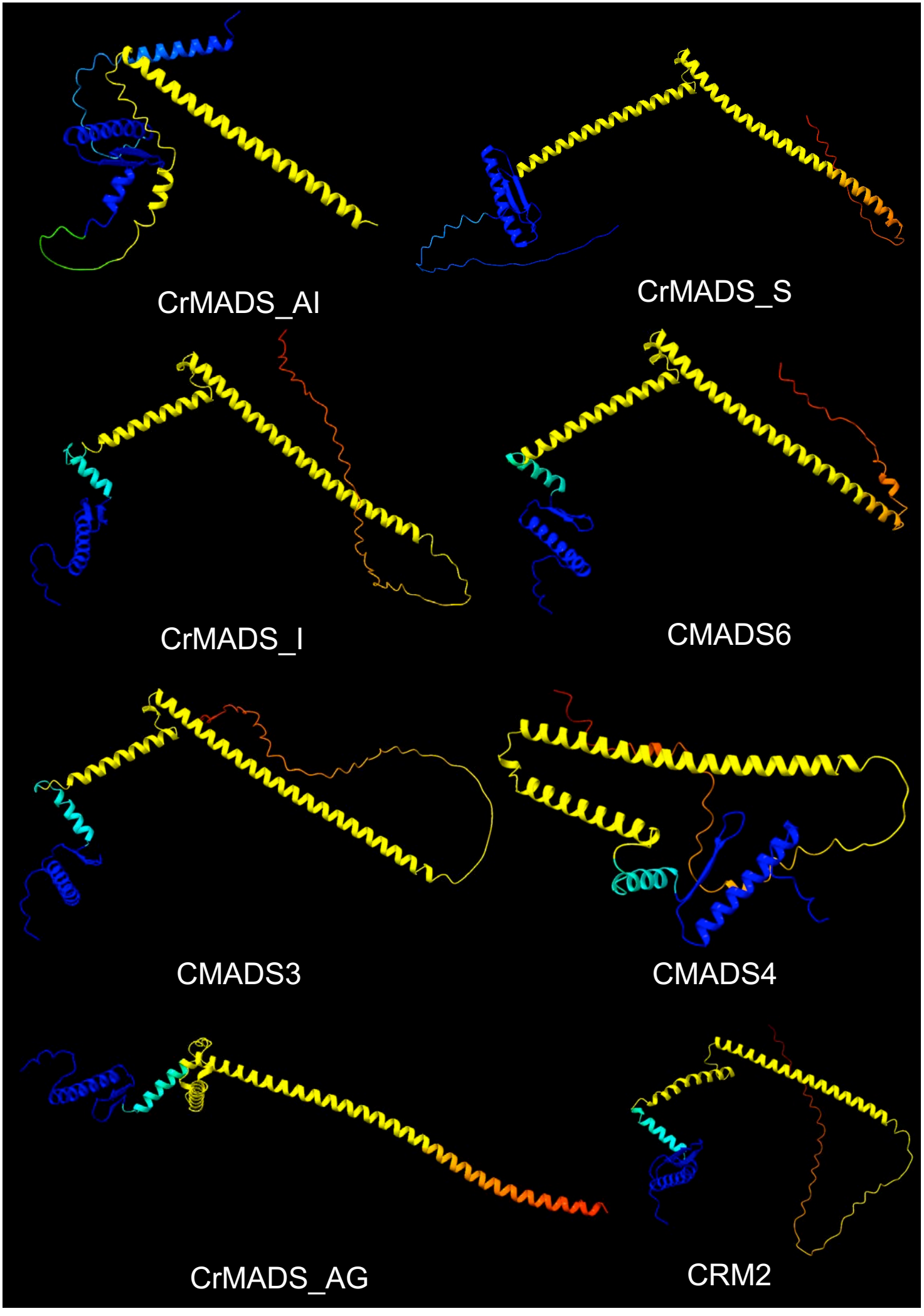

CrMADS_AI
CrMADS_S
CrMADS_I
CMADS6
CMADS3
CMADS4
CrMADS_AG
CRM2

Figure 5.8: Protein structural modelling of all identified Type-II MIKC^c MADS-box proteins in *Ceratopteris*. Structures were predicted using AlphaFold2 run on ColabFold and visualised using ChimeraX (UCSF ChimeraX, 2023). The blue sections represent the M domain, the yellow sections represent the K domain.

The phylogenetic trees presented in this chapter (Figures 5.4 and 5.6) show the clade containing *Ceratopteris* proteins CMADS6 and CrMADS_G as closely related to the AG clade. With the CMADS6 and AG clades forming one clade when the highly conserved M domains have their relativity determined, this. Previously the candidate protein CMADS1 had been identified as the most related (Hasebe *et al*., 1998). A comparison between the predicted structures of CMADS1, CMADS6, and CrMADS_G with the putative *Arabidopsis* orthologue, AG was presented in Figure 5.9 to observe the structural similarities between them.

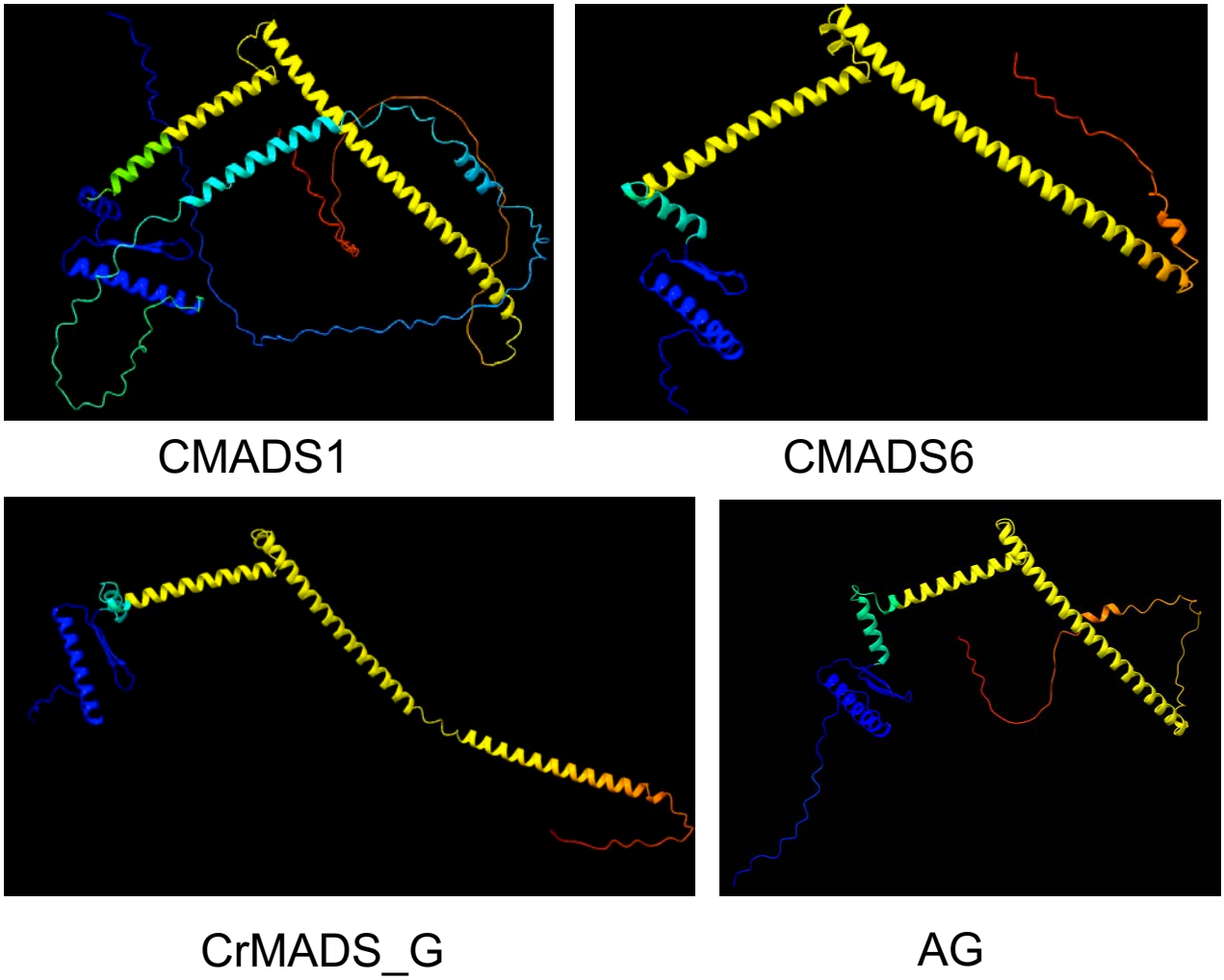


Figure 5.9: Protein structural modelling of CMADS1, CMADS6, CrMADS_G, and AG. Structures were predicted using AlphaFold2 run on ColabFold and visualised using ChimeraX (UCSF ChimeraX, 2023). The blue sections represent the M domain, the yellow sections represent the K domain.

5.5. Yeast-II-Hybrid assay to determine protein-protein interactions between *Ceratopteris* proteins.

To gain further information about potential functions of *Ceratopteris* MADS box proteins, interactions between selected proteins were investigated using yeast-II-hybrid assays.

Yeast-II-Hybrid (Y2H) is often used to observe if there are interactions between proteins. This is achieved as the system is set up so that expression of a. reporter gene depends on the interaction of a pair of transcription factors or other proteins. In the Y2H system a chosen protein pair are made into fusion proteins that contain both halves of a yeast transcription factor needed for reporter gene expression as well as the test protein (Figure 2.6), however this system is prone to false positives (Brückner *et al*., 2009). To increase accuracy of the library, a bait-dependency test can be performed (Möckli *et al*., 2007). Identified preys and unrelated baits are tested against one another, which can eliminate false positives resulting from non-specific interactions with the bait or other “sticky” interactions over coming selection.

As MADS-box proteins are transcription factors, they can offer autoactivation within the system that can lead to issues during Y2H, with self-activation causing false positives. Conserved motifs were found in the carboxy-terminal domain of MADS-box proteins, and transcriptional activation capacity has been detected in the C-terminal domain (Cho *et al*., 1999; Goto *et al*., 2001). To counteract this, truncated forms of the proteins were used, with the proteins truncated after the last predicated alpha helix of

the K domain as transcriptional activity has been attributed to the C-domain. However, MADS-box proteins in *Arabidopsis* have also been successfully used in their entirety in Y2H (de Folter *et al*., 2005). It has been shown that truncated proteins and full proteins show the same interaction results (Davies *et al*.,1996).

The floral quartet model suggests a quaternary structure of tetramers for proteins specifying the different floral organs (Theißen and Saedler, 2001). However, interactions have still been observed between pair-wise assays using Yeast-II-Hybrid (de Folter *et al*., 2005). As such, Yeast-II-Hybrid assays were performed between CMADS1 and the two other proteins that fall within its phylogenetic clade *Cr*MADS_H and *Cr*MADS_S, according to the *Ceratopteris* MADS-box Type-II MIKC^{c} phylogenies in Chapter 3 and this chapter (Figures 3.2, 5.4-5.6). Moreover, the genes encoding the proteins show similar expression patterns in reproductive tissue (Figure 3.5). Within these assays, all proteins also had their pairwise self-interactions analysed as homodimerisation is a common occurrence within the MADS-box family (Honma and Goto, 2001).

*At*MYB93 and *At*MPK3 were chosen as positive controls (Cao and Coates, unpublished). Initially, full-length sequences were cloned for interaction analysis, however the C domain of MADS-box proteins has previously caused false positives due to its DNA transactivation and complex formation properties (Lai *et al*., 2021). Thus, sequences were truncated after the K domain ended. Whether this truncation affects interaction is yet to be determined. A full-length CMADS1 protein was used to act as a comparison, but unfortunately the full-length sequence could not be cloned into the prey vector pGBKT7 during the timeframe of this analysis. The full-length

sequence assay performed with CMADS1 in the bait vector pGADT7 had poor growth on SD-LT plates that contain all necessary nutrients for healthy growth suggesting that although insertion was achieved the yeast was not able to thrive.

Yeast were transformed with relevant pairs of bait/prey constructs, namely protein plus empty vector to test for autoactivation, or two proteins to test for interaction (Figure 5.10-5.12, representing 3 independent rounds of yeast transformation). Initially, transformed yeast was grown on dropout medium lacking Leucine and Tryptophan to ensure yeast transformation with both bait and prey constructs was successful (Galletta and Rusan, 2015) (Figures 5.10-5.12). Quadruple dropout medium lacking adenine, histidine, leucine, and tryptophan was used as a stringent test to show interactions between proteins (Shivhare *et al*., 2021). A triple dropout of adenine, leucine, and tryptophan was used as a less stringent control to observe any weak interaction between proteins (Figure 5.10-5.12). As no growth was observed on the triple dropout another medium was used with a dropout of histidine, leucine, and tryptophan, however this dropout media is known to be “leaky”, so the addition of 3-AT was used to prevent false positives (Caufield, Sakhawalkar and Uetz, 2012) using different 3-AT concentrations to observe any potential weaker interactions present (Figure 5.10-5.12).

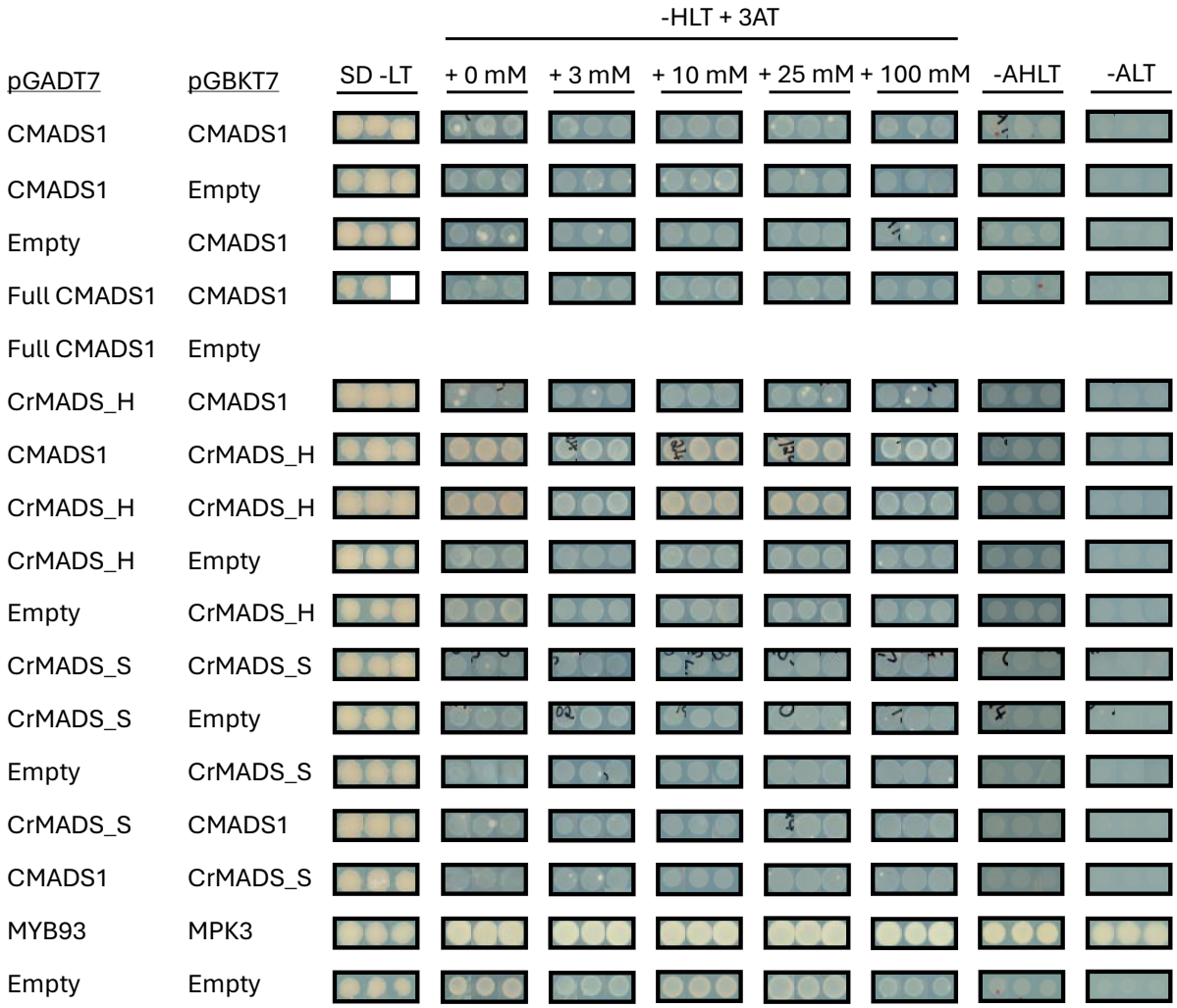


Figure 5.10: One biological replicate of Yeast-II-Hybrid interactions after 5 days of growth at 30 °C. *AtMYB93* and *AtMPK3* act as positive control whilst empty vectors act as negative controls.

Different stringencies of media were used to ensure that even weak level interactions were observed. However, no robust interactions were seen between any MADS box protein combination tested when compared to both positive and negative controls in any of the three biological repeats (Figure 5.10-5.12). Although there is some growth on -HLT seen for both CMADS1 homodimerisation and CMADS_H homodimerization, the negative controls also show similar low levels of growth and all growth is less than for the positive controls (Figure 5.10-5.12).

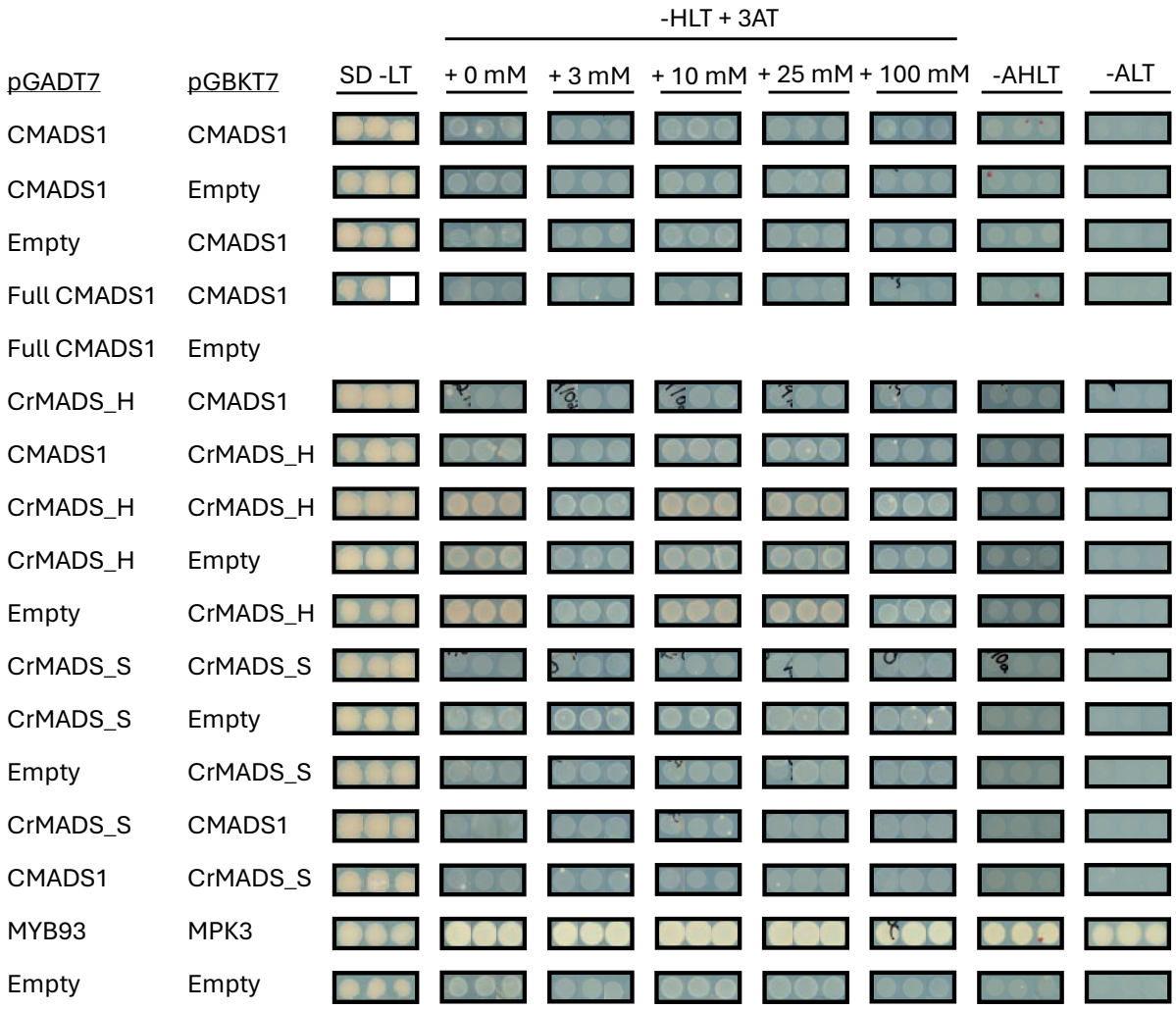


Figure 5.11: A second biological replicate of Yeast-II-Hybrid interactions after 5 days of growth at 30 °C. *AtMYB93* and *AtMPK3* act as positive control whilst empty vectors act as negative controls.

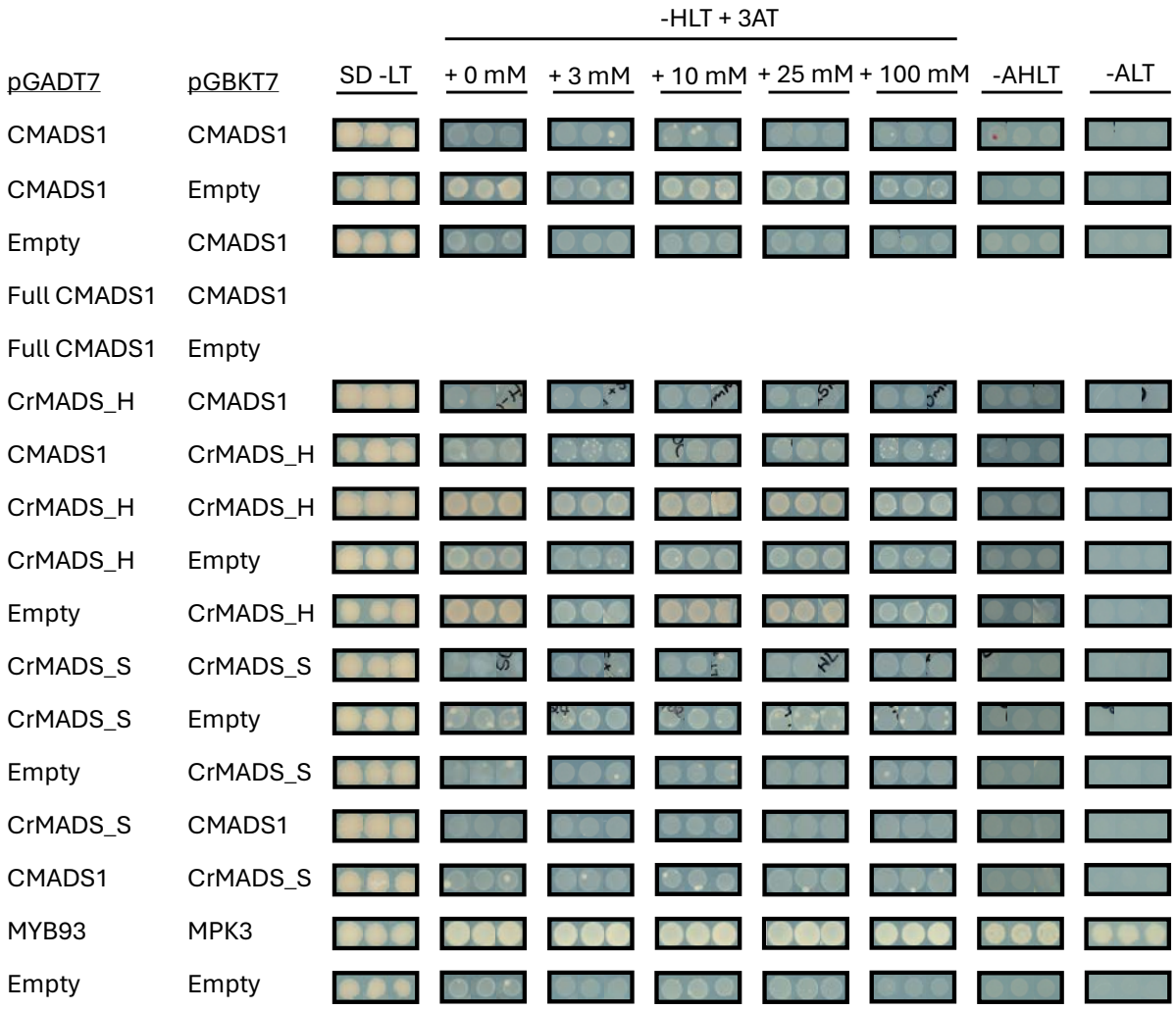


Figure 5.12: A third biological replicate of Yeast-II-Hybrid interactions after 5 days of growth at 30 °C. Only two colonies grew in Full CMADS1 interaction assay so no third biological replicate could be performed. *AtMYB93* and *AtMPK3* act as positive control whilst empty vectors act as negative controls.

5.6. Discussion

5.6.1. There is high sequence similarity between the conserved domains of MADS-box Type-II MIKC^c proteins in *Ceratopteris* and *Arabidopsis* but some divergence of predicted structure.

The 14 subfamilies identified in the phylogenetic tree of *Arabidopsis* and *Ceratopteris* Type-II MIKC^c MADS-box proteins with all domains present are reflected in previous phylogenetic trees that contained fewer proteins (Hasebe *et al*., 1998; Munster *et al*., 1997). Twelve *Arabidopsis* monophyletic clades were identified in the maximum likelihood phylogenetic models, these being categorised as: AGL6, ABS, SEP, AP1, SOC1, AP3/PI, SVP, AGL15, ANR1, AG, FUL, and FLC (Figure 5.4). This is concurrent with previous *Arabidopsis* phylogenetic analysis (Becker and Theissen, 2003). Within the *Ceratopteris* clade, CMADS3 contains the most proteins of the three *Ceratopteris* clades and is also the least related, diverging further back in deep evolutionary history when compared to CMADS1 and CMADS6 clades (Reeves and Richards, 2007). CMADS1 and proteins that fall within the CMADS1 clade (Figure 3.2) show an extended disordered region at the N-terminus, this shows a higher similarity with AGAMOUS (Mizukami *et al*., 1996).

The MADS-box family is present amongst most Eukaryotes. Despite being present in a great range of species, certain sequence domains are highly conserved. For example, the Type-II MADS-box proteins that are present in plants have an animal counterpart subcategory of MEF2-like MADS-box proteins (Gramzow and Theissen, 2010; Qui *et al*., 2023). Despite differing functions, these proteins contain the

conserved motifs of both the M-domain and the K-domain secondary structures presented in this chapter (Smaczniak *et al*., 2012).

The well-researched functions of the MADS-box family in angiosperms, here represented by the model organism *A. thaliana*, suggest a high similarity between structures (Abdullah-Zawawi *et al*., 2021). This is supported by the predicted secondary structures that can be observed on the protein database Uniprot (Bateman *et al*., 2020) and within this chapter with the highly conserved M and K domains highlighted. In the interest of comparison, *Ceratopteris* MADS-box Type-II MIKC^{c} proteins had their secondary structures predicted using AlphaFold2 (UCSF ChimeraX, 2023) as their function has only been hypothesised (Münster *et al*., 1997) Predicting their structure could offer insight into their functions in *Ceratopteris* and potentially non-seed plants more generally.

Predictions of proteins secondary structures rely on models that can be more or less suited to the protein depending upon their sequence. For example, proteins with high conservation will need a different model to those with low conservation (Whisstock and Lesk, 2003). The models used in this chapter were chosen due to the small number of sequences being aligned with highly conserved structural motifs (E-INS-i model chosen for those with multiple conserved domains, L-INS-i chosen for the M-domain only alignment). For sequences where low conservation is expected, another model such as the G-INS-i model would be a better fit (Huang *et al*., 2023). Proteins with repeat motifs will need a different model to those with more disordered regions (Lotthammer *et al*., 2024).

Ceratopteris MADS-box proteins had not had their secondary structure predicted in the past, instead the high conservation observed through the family across different species was trusted to encompass secondary structure similarity. The work in this chapter shows that further study into the secondary structures of the MADS-box family is necessary as the predictions alone suggest a higher differentiation in *Ceratopteris* when compared to *Arabidopsis*. Although ChimeraX uses an AlphaFold server to predict protein secondary structures, validation of these predications is essential. Structural studies such as Nuclear Magnetic Resonance Spectroscopy and X-ray diffraction could lead to more accurate protein structure prediction (Alberts *et al*., 2002). Understanding the secondary structures can help determine the functions of these proteins within non-seed and non-flowering plants.

5.6.2. Relatedness of Ceratopteris and Arabidopsis MADS-box Type-II MIKC^c proteins.

Type-II MIKC^c proteins contain four domains of differing conservation. The M-domain is the highest conserved with a function of DNA-binding, K-domain is strongly conserved but not as much as the M-domain with a function in protein dimerisation, I-domain is somewhat conserved with no recurring motifs with an hypothesised function in the determination of DNA-binding dimers, and the C-domain is the least conserved often diverging quite largely despite holding a transcriptional activation function and often being involved in the formation of multimeric complexes (Becker and Theissen, 2003).

Three phylogenetic trees were created to allow for comparison of relatedness when compared with domain conservation. The lowest bootstrap value is seen in the M-domain only phylogeny which would be expected to have the highest level of relatedness and certainty, this could suggest that all parts of the protein are important for robust alignments with the larger divergence representing more accurate relationships.

5.6.3. There is greater variability between *Ceratopteris* MADS-box Type-II MIKC^c protein secondary structure than that observed in *Arabidopsis*.

The limited crystal structures of MADS-box proteins in plants have led to the reliance on predicted secondary structures. However, much can be determined about protein function from their secondary structure, for example, the MADS-domain is able to function as a DNA binding domain through the β-strands of two MADS-box proteins interacting and forming β-sheets (Gramzow and Theissen, 2010). The increase in the accessibility of sequences and predicted protein structures on databases such as Uniprot (Bateman *et al*., 2022) and The Protein Data Bank (Berman, 2000) has led to a gap in the sequence:function knowledge of many proteins (Gligorijević *et al*., 2021).

This chapter sought to generate predicted secondary structures of *Ceratopteris* MIKC^c MADS-box proteins using ChimeraX and the AlphaFold2 prediction extension run on Google Colab (Mirdita *et al*., 2022). *Arabidopsis* MIKC^c MADS-box proteins also had their secondary structures predicted using the same method to allow for better

comparison between the two species. The predicted structures for the *Arabidopsis* proteins reflect what is expected, with high conservation between the structures (Parenicova *et al*., 2003). However, the *Ceratopteris* predicted structures show a more divergence. This could suggest more diverse functions, as the *Arabidopsis* proteins are categorised into the ABCDE model due to their function in floral organogenesis, the lack of floral organs in *Ceratopteris* could mean the MADS-box proteins hold the more diverse functions seen in ancestral land plants such as the early cell development functions held by Bryophyte MADS-box proteins (Thangavel and Nayar, 2018).

Secondary structure prediction currently relies upon experimental validation and functional studies to determine protein functions; computational methods that predict function based on structure rely on comparison to known protein networks (Mostafavi *et al*., 2008). This limits the use for lesser studied species or gene families. However, computational methods to determine protein function solely relying on sequence prediction would allow for prediction of function of proteins in unknown organisms and species without the required network data (Gligorijević *et al*., 2021).

5.6.4. No interactions were observed between proteins in the *CMADS1* clade using Yeast-II-Hybrid assays.

Protein-protein interactions between MADS-box proteins have been observed using Yeast-II-Hybrid in many angiosperm species including *Arabidopsis* (de Folter *et al*., 2005). It is hypothesised that MADS-box proteins in ancestral land plants such as the bryophyte *Physcomitrium* interact in complex networks similar to those present in

flowering plants (Henschel *et al*., 2002). Despite MADS-box proteins forming tetramers within the floral quartet model (Theißen, Melzer and Rümpler, 2016), dimerisation has also been observed between proteins within the family (West *et al*., 1998). As such Yeast-II-Hybrid is an appropriate assay for determining protein-protein interactions within the MADS-box protein family.

Before this work, *Ceratopteris* MADS-box proteins had yet to have any interactions investigated. This chapter used Yeast-II-Hybrid to observe any binary interactions between CMADS1 and the two proteins that fall within its clade (Figure 5.10-5.12) *Cr*MADS_H and *Cr*MADS_S. To avoid auto-activation, proteins were truncated at the K-domain to remove the C-terminus that has previously been shown to cause autoactivation (Immink *et al*., 2002). A full-length CMADS1 protein was successfully cloned into the prey vector pGADT7 for comparison potential interactions between truncated and full-length proteins. No interactions were reported in this assay. Future work could continue the Yeast-II-Hybrid assay begun in this chapter to encompass other clades of the *Ceratopteris* MIKC^{c} proteins to potentially identify protein interactions outside of the closely related proteins. Alternatively future work could utilise other *in vitro* methods such as pull-down assays (Louche, Salcedo and Bigot, 2017) or co-immunoprecipitation (Tan and Yammani, 2022). *In vivo* methods such as Bimolecular Fluorescence Complementation (BiFC) could also be performed although this has yet to be achieved in *Ceratopteris* (Ohad, Shichrur and Yalovsky, 2007).

Further investigations into *Ceratopteris* MADS-box protein function could also investigate protein-DNA interactions as in *Arabidopsis* (Aerts *et al*., 2018) either via large-scale ChIP-seq screens, or bioinformatic analysis of CArG-containing promoters

in *Ceratopteris* followed by targeted ChIP-PCR of hypothesised target genes. Putative targets could be those with similar secondary structures to AG for example the CMADS1 clade. Alternatively, relatedness to AG could be used with CMADS6 and CrMADS_G that fall within the AG clade being putative targets (Figures 5.4-5.6).

CHAPTER 6: DISCUSSION

6.1. Introduction

MADS-box genes are present from charophytes to angiosperms and are best documented in angiosperms (Thangavel and Nayar., 2018). The sister clade to spermatophytes, the monilophytes, has remained largely unexplored. With the full genome of the model fern *Ceratopteris* not being sequenced until 2019 and updated in 2022 (Marchant *et al*., 2019; 2022) it was hard to determine at a molecular level the reproductive evolutionary changes (Schoe *et al*., 2019) that occurred within this clade to enable plants to transition from a free-living gametophyte stage to the enclosed seed reproductive system.

CMADS1 is a transcription factor within the MADS-box family in *Ceratopteris* that was previously hypothesised to be an *AGAMOUS* homologue (Hasebe *et al*., 1997). This project sought to determine the similarities between the two transcription factors. To achieve this, both *in vivo* and *in vitro* approaches were taken to address the following aims: (i) Defining the full *Ceratopteris* MADS-box family and categorising the genes within both phylogenetically and molecularly (ii) Investigating the function of *CMADS1 in vivo* through transgenic approaches in *Ceratopteris* and exogenous overexpression of *CMADS1* in *Arabidopsis* when compared to overexpression of native *AGAMOUS* (iii) Understanding *CMADS1* protein function via *in silico in vitro* approaches, using protein modelling, bioinformatics, and Yeast-II-Hybrid assays.

6.2. *Ceratopteris* as a tool for closing the gap in reproductive evolutionary knowledge

Despite monilophytes holding a position that puts them as the sister clade to all seed plants, at a pinnacle point when plants were switching from a free-living gametophyte life stage to a much smaller gametophyte stage, the monilophyte MADS-box transcription factor family and their functions have yet to be fully determined. The MADS-box family is known in seed and flowering plants to be responsible for reproductive organ development (Aerts *et al*., 2018) and known in ancestral reproductive systems to be responsible for overall early cell development (Alvarez-Buylla *et al*., 2000). However, MADS-box gene function in monilophytes remains unknown. *Ceratopteris* is a fern within the monilophyte clade that has been successfully transformed previously (Plackett *et al*., 2014). Both gametophytes and sporophytes can be grown in tissue culture and do not exhibit negative effects due to being grown in a laboratory setting (Kinosian and Wolf, 2022). This makes *Ceratopteris* a good candidate for use in determining the functions of MADS-box proteins within the monilophyte clade. This combined with the published genome (Marchant *et al*., 2019; Marchant *et al*., 2022) allows for research both *in vivo* and *in vitro* with experimental assays possible at a molecular level unlike in other ferns.

The complete *Ceratopteris* genome has allowed this project to conclude that *Ceratopteris* MADS-box proteins fall into 3 clades, suggesting gene duplication of the MADS-box family had occurred prior to the divergence of ferns (Chapter 3). Expression analysis of *Ceratoperis* MADS-box genes determined MADS-box genes

with differing patterns of expression and those enriched in reproductive tissue (Chapter 3). Following on from this, transgenic plants were then produced with the intention of knocking down *CMADS1* expression in *Ceratopteris* (Chapter 4). Future work can build off this base of cloning and transformation that has generated valuable molecular resources, by taking the lowest-expressing *CMADS1* lines generated through to homozygosity for further phenotyping assays. The constructs for over-expression of both myc-tagged and untagged *CMADS1* in *Ceratopteris* can be used to transform *Ceratopteris* and generate homozygous lines for phenotyping. The transformed lines with potential over-expression of *CMADS1* and *AG* in *Arabidopsis* can be grown on further to homozygosity, checked for over-expression via RT-PCR and (in the case of myc-tagged proteins) western blotting, and assayed for reproductive phenotypes.

6.3. MADS-box family and its vast role in reproductive evolution

The MADS-box family holds multiple functions throughout multiple species from broad homogeneous developmental functions to those specific to floral organ identity. Within charophytes, the sister group to extant land plants, few MADS-box genes have been identified, with only Type-II MIKC^c being published (Gramzow *et al*., 2023). Expression analysis suggests that charophyte MADS-box genes function in haploid reproductive cell development and differentiation (Tanabe *et al*., 2005). As the MADS-box family evolved along with the complexity of land plants, more sub-categories developed with gene duplications leading to the subcategories of Type-I and Type-II MADS-box genes (described in chapter 1). Prior to the advent of bryophytes the subcategories within Type-II of MIKC^c and MIKC* evolved (Henschel *et al*., 2002). However, bryophyte

MADS-box Type-I and Type-II present different clustering suggesting angiosperm MIKC orthologues are not found in mosses (Thangavel and Naya, 2018). The MIKC^c family is highly categorised as it contains the transcription factors responsible for floral organ identity, floral timing, and fruit development in angiosperms (Lovisetto *et al*., 2011; Gramzow *et al*., 2023).

The monilophyte clade, the sister clade to spermatophytes, provides a way to compare two different plant reproductive systems: a system that relies on spores and free-living gametophytes versus a system that uses seeds, which develop after the presence of reduced gametophyte that is fully dependent on surrounding maternal sporophytic tissue. This project suggests that both the function of early cell development and the more specialised reproductive functions could be held by the MADS-box family proteins within the model fern *Ceratopteris*. This is due to the differences in developmental expression patterns shown, with two genes remaining consistently present through different developmental tissues, this number could be higher however this study was limited as the focus remained on reproductive functions so not all genes had their expression in gametophyte and vegetative tissue checked. A smaller subset of genes, specifically those that fall within the CMADS1 clade, showed differing expression with enriched expression in the reproductive sporophyte tissue. Moreover, *Ceratopteris* MADS-box proteins show differences in predicted protein structures that are not observed in *Arabidopsis* MIKC^c MADS-box proteins.

6.4. *CMADS1* as a potential homolog of *AGAMOUS*

It is shown in Chapter 5 of this project that *CMADS1* and the associated clade have similarities with the *Arabidopsis* MADS-box protein *AGAMOUS* through an elongated disordered region at the N-terminus. It has been stated that the clade CRM6, called in this project *CMADS1* and the within *Ceratopteris* MADS-box proteins within this clade could hold a closer relation to *AGAMOUS* due to where they sit phylogenetically (Hasebe *et al*., 1997). This combined with the differential expression observed across *Ceratopteris* developmental stages, with increased expression in reproductive tissues when compared with both gametophyte and vegetative tissues (Figure 3.5) suggest a close relationship in function to *AGAMOUS* when compared with other *Ceratopteris* MADS-box genes that show a more homogeneous expression. These homogeneously expressed MADS-box genes may hold a more ancestral role in *Ceratopteris,* reflecting the broader, and perhaps redundant, functions of MADS-box genes in both reproductive and non-reproductive tissue in bryophytes (Tanabe *et al*., 2018).

6.5. Future work

This project has identified MADS-box genes with differences in expression across developmental tissues using RT-PCR and qPCR. Expanding this work to include additional methods would resolve the observed discrepancies allowing for a greater understanding of how the expression patterns of the MADS-box genes affect their functions within *Ceratopteris.* Establishing the expression patterns of the *Ceratopteris* MADS-box family through northern blotting of a full range of tissues would build on the

expression analysis presented in Chapter 3 of this project that focussed on different developmental stages of largely frond tissue only. Focussing on different tissues of *Ceratopteris* could provide insight into the functions of these proteins within the monilophytes life cycle. Using northern blotting, or even *in situ* hybridisation to determine expression patterns instead of RT-PCR and qPCR would allow for the discrepancies observed between RT-PCR gene expression patterns and qPCR expression patterns to be removed as there is not any amplification bias (Dean, Goodwin and Hsiang, 2002). Alongside this, tissue and stage specific RNA-seq could offer an alternative method of identifying gene expression patterns in quantitative detail (Corchete et al., 2020). These methods would help resolve discrepancies between the RT-PCR and q-PCR data as well as offer long-term resources for *Ceratopteris*. Available RNA-seq data (McCready, unpublished) shows an increase of expression of the CMADS1 clade in reproductive tissue, reflecting the RT-PCR data presented in this thesis.

Research into *Ceratopteris* transformation efficiency such as the construction of smaller delivery vectors to allow for higher plasmid yields and more reliable cloning (Pratt, Knoblauch and Kunz, 2020). The use of double-barrelled biolistic guns in the bombardment process could lead to a more reliable method of fern transformation and allow for work to be achieved in this clade of plants that was not previously possible (Miller *et al*., 2021). As such plants could be taken through more generations as the cloning and transformation stage time frame could be reduced due to increased reliability and efficiency. Alternatively, although this project used particle bombardment for transformation of *Ceratopteris*, CRISPR/Cas9 methods have since been developed (Xiang and Li, 2024) that allow a much shorter time span of transformation by using

gametophytes rather than induction of sporophytic callus (Plackett *et al*., 2014). With the advent of CRISPR/Cas9 gene editing and over expression being viable in *Ceratopteris* future work could consist of knock out mutants as only knock down was possible thus far. In addition, over-expression transformants could utilise different promotors and transformants could be more accurately determined through optimised *Agrobacterium* transformations (Jiang *et al*., 2024).

This project shows no evidence of protein-protein interactions between CMADS1 and itself or other proteins within its phylogenetic clade, *Cr*MADS_S and *Cr*MADS_H. In future research it would be best to concentrate on whether *Ceratopteris* MADS-box proteins in other clades interact with those from CMADS1 clade as protein-protein interaction is an expected characteristic of the MADS-box family (Immink *et al*., 2002). Yeast-3-Hybrid could be used to observe protein-protein interactions of higher-order complexes than can be observed in the binary interactions reported by Y2H (Glass and Takenaka, 2018). This would be a valid next step due to the complexes that have been characterised in the floral quartet (Theißen and Saedler, 2001). *In planta* approaches such as co-immunoprecipitation could be used to validate Y2H results (Burckhardt, Minna and Danuser, 2021).

6.6. Final conclusions

Through amplification, cloning, and sequencing 33 *Ceratopteris* MADS-box genes were determined to have unique coding sequences. Of these MADS-box genes 17 were determined to be Type-II and 14 Type-I, two remain uncategorised. The most comprehensive phylogenetic analysis to date of MADS-box proteins in *Ceratopteris*

was produced in Chapter 3 with a multi-species phylogenetic tree with representative species from each land plant group. Comparison phylogenetic trees of *Ceratopteris* and *Arabidopsis* MADS-box proteins, along with phylogenetic trees of truncated MADS-box proteins to show relationships between proteins dependent on the conservation of the domains were produced in Chapter 4. In depth expression analysis of fern MADS-box homologues was produced in Chapter 3, with focus on Type-II MIKC^{c} genes. Semi-quantitative RT-PCR showed an increase in expression of MADS-box genes within the CMADS1 clade specifically in reproductive tissue, suggesting a regulated expression pattern. Although qRT-PCR differed from RT-PCR in the relative expression levels reported there was still differential expression of the CMADS1 clade genes in different developmental stages of *Ceratopteris.* This suggests *Ceratopteris* MADS-box genes have specific expression as is seen is angiosperm MADS-box genes, rather than homogeneous expression as is observed in ancestral plants such as the bryophytes (Zhong *et al*., 2024; Qiu, Li and Köhler, 2024).

Chapter 4 developed resources for making transgenic ferns that overexpress and knockdown *CMADS1* in addition to molecular resources for over expression of CMADS*1* in *Ceratopteris* and *Arabidopsis.* Chapter 5 characterised conserved protein domains of all *Ceratopteris* MADS-box proteins and used this knowledge to model predicted structures all of MIKC^{c} *Ceratopteris* MADS-box proteins. Protein-protein interactions of the CRM6 clade of MADS-box proteins present in *Ceratopteris* were analysed through Yeast-II-Hybrid assays and no interactions were detected.

Overall, the work in this project has produced resources for future work with the MADS-box family in *Ceratopteris* providing a vital link between spore and seed based reproductive models.

REFERENCES

Abdullah-Zawawi, M.-R., Ahmad-Nizammuddin, N.-F., Govender, N., Harun, S., Mohd-Assaad, N. and Mohamed-Hussein, Z.-A. (2021). Comparative genome-wide analysis of WRKY, MADS-box and MYB transcription factor families in Arabidopsis and rice. *Scientific Reports*, 11(1). doi:https://doi.org/10.1038/s41598-021-99206-y.

Aerts, N., de Bruijn, S., van Mourik, H., Angenent, G.C. and van Dijk, A.D.J. (2018). Comparative analysis of binding patterns of MADS-domain proteins in Arabidopsis thaliana. *BMC Plant Biology*, 18(1). doi:https://doi.org/10.1186/s12870-018-1348-8.

Aggarwal, P., Krishna Reddy Challa, Rath, M., Sunkara, P. and Nath, U. (2018). Generation of Inducible Transgenic Lines of Arabidopsis Transcription Factors Regulated by MicroRNAs. *Methods in molecular biology*, pp.61–79. doi:https://doi.org/10.1007/978-1-4939-8657-6_4.

Agrawal, N., Dasaradhi, P.V.N., Mohmmed, A., Malhotra, P., Bhatnagar, R.K. and Mukherjee, S.K. (2003). RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews*, 67(4), pp.657–685. doi:https://doi.org/10.1128/mmbr.67.4.657-685.2003.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). *Molecular Biology of the Cell*. 4th ed. New York: Garland Science.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martinez-Castilla, L. and Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences*, 97(10), pp.5328–5333. doi:https://doi.org/10.1073/pnas.97.10.5328.

arabidopsis.info. (n.d.). *Eurasian Arabidopsis Stock Centre (uNASC)*. [online] Available at: https://arabidopsis.info/.

Banks, J.A. (2018). Fern genomes finally here. *Nature Plants*, 4(7), pp.404–405. doi:https://doi.org/10.1038/s41477-018-0202-1.

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L.G. and Garmiri, P. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), pp.D480–D489. doi:https://doi.org/10.1093/nar/gkaa1100.

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L.J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A. and Insana, G. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1). doi:https://doi.org/10.1093/nar/gkac1052.

Becker, A. and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution*, 29(3), pp.464–489. doi:https://doi.org/10.1016/s1055-7903(03)00207-0.

Becker, B. (2013). Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends in Plant Science*, 18(4), pp.180–183. doi:https://doi.org/10.1016/j.tplants.2012.09.010.

Berman, H.M. (2000). The Protein Data Bank. *Nucleic Acids Research*, [online] 28(1), pp.235–242. doi:https://doi.org/10.1093/nar/28.1.235.

Blomenkemper, P., Kerp, H., Hamad, A.A., DiMichele, W.A. and Bomfleur, B. (2018). A hidden cradle of plant evolution in Permian tropical lowlands. *Science*, [online] 362(6421), pp.1414–1416. doi:https://doi.org/10.1126/science.aau4061.

Boc, A. (2024). *Trex-online*. [online] Uqam.ca. Available at: http://www.trex.uqam.ca/index.php?action=mafft [Accessed 26 Sep. 2024].

Bowles, A.M.C., Williamson, C.J., Williams, T.A., Lenton, T.M. and Donoghue, P.C.J. (2022). The origin and early evolution of plants. *Trends in Plant Science*. doi:https://doi.org/10.1016/j.tplants.2022.09.009.

Bowman, J.L. (2022). The liverwort Marchantia polymorpha, a model for all ages. *Current Topics in Developmental Biology*, pp.1–32. doi:https://doi.org/10.1016/bs.ctdb.2021.12.009.

Bowman, J.L. and Edwige Moyroud (2024). Reflections on the ABC model of flower development. *The Plant Cell*. doi:https://doi.org/10.1093/plcell/koae044.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S.S., Althoff, F., Araki, T., Arteaga-Vazquez, M.A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C.R. and Briginshaw, L. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell*, [online] 171(2), pp.287-304.e15. doi:https://doi.org/10.1016/j.cell.2017.09.030.

Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (2012). The ABC model of flower development: then and now. *Development*, 139(22), pp.4095–4098. doi:https://doi.org/10.1242/dev.083972.

Brückner, A., Polge, C., Lentze, N., Auerbach, D. and Schlattner, U. (2009). Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *International Journal of Molecular Sciences*, 10(6), pp.2763–2788. doi:https://doi.org/10.3390/ijms10062763.

Bui, L.T., Cordle, A.R., Irish, E.E. and Cheng, C.-L. (2015). Transient and stable transformation of Ceratopteris richardii gametophytes. *BMC Research Notes*, [online] 8(1). doi:https://doi.org/10.1186/s13104-015-1193-x.

Bui, L.T., Pandzic, D., Youngstrom, C.E., Wallace, S., Irish, E.E., Szövényi, P. and Cheng, C.-L. (2017). A fern*AINTEGUMENTA*gene mirrors*BABY BOOM*in promoting apogamy in*Ceratopteris richardii*. *The Plant Journal*, 90(1), pp.122–132. doi:https://doi.org/10.1111/tpj.13479.

Burckhardt, C.J., Minna, J.D. and Danuser, G. (2021). Co-immunoprecipitation and semi-quantitative immunoblotting for the analysis of protein-protein interactions. *STAR Protocols*, 2(3), p.100644. doi:https://doi.org/10.1016/j.xpro.2021.100644.

C-Fern. (n.d.). *C-Fern*. [online] Available at: https://c-fern.org/.

Callens, C., Tucker, M.R., Zhang, D. and Wilson, Z.A. (2018). Dissecting the role of MADS-box genes in monocot floral development and diversity. *Journal of Experimental Botany*, 69(10), pp.2435–2459. doi:https://doi.org/10.1093/jxb/ery086.

Caufield, J.H., Sakhawalkar, N. and Uetz, P. (2012). A comparison and optimization of yeast two-hybrid systems. *Methods*, 58(4), pp.317–324. doi:https://doi.org/10.1016/j.ymeth.2012.12.001.

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), pp.1255–1265. doi:https://doi.org/10.1534/genetics.115.182964.

Chen, F., Zhang, X., Liu, X. and Zhang, L. (2017). Evolutionary Analysis of MIKCc-Type MADS-Box Genes in Gymnosperms and Angiosperms. *Frontiers in plant science*, 8. doi:https://doi.org/10.3389/fpls.2017.00895.

Chen, R.-G., Ma, J., Luo, D., Hou, X., Ma, F., Zhang, Y., Meng, Y., Zhang, H. and Guo, W. (2019). CaMADS, a MADS-box transcription factor from pepper, plays an important role in the response to cold, salt, and osmotic stress. 280, pp.164–174. doi:https://doi.org/10.1016/j.plantsci.2018.11.020.

Cho, S., Jang, S., Chae, S., Chung, K.M., Moon, Y.-H., An, G. and Jang, S.K. (1999). Analysis of the C-terminal region of Arabidopsis thaliana APETALA1 as a transcription activation domain. *Plant Molecular Biology*, 40(3), pp.419–429. doi:https://doi.org/10.1023/a:1006273127067.

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. *The Plant Journal*, 16(6), pp.735–743. doi:https://doi.org/10.1046/j.1365-313x.1998.00343.x.

Coen, E.S. and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature*, 353(6339), pp.31–37. doi:https://doi.org/10.1038/353031a0.

Corchete, L.A., Rojas, E.A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N.C. and Burguillo, F.J. (2020). Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Scientific Reports*, 10(1). doi:https://doi.org/10.1038/s41598-020-76881-x.

D'Apice, G., Moschin, S., Nigris, S., Ciarle, R., Muto, A., Bruno, L. and Baldan, B. (2022). Identification of key regulatory genes involved in the sporophyte and gametophyte development in *Ginkgo biloba* ovules revealed by in situ expression analyses. *American Journal of Botany*, 109(6), pp.887–898. doi:https://doi.org/10.1002/ajb2.1862.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B. and Flouri, T. (2019). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*. doi:https://doi.org/10.1093/molbev/msz189.

Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H. and Sommer, H. (1996). Multiple interactions amongst floral homeotic MADS box proteins. *The EMBO Journal*, 15(16), pp.4330–4343. doi:https://doi.org/10.1002/j.1460-2075.1996.tb00807.x.

de Folter, S., Immink, R.G.H., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B. and Angenent, G.C. (2005). Comprehensive Interaction Map of the Arabidopsis MADS Box Transcription Factors. *THE PLANT CELL ONLINE*, 17(5), pp.1424–1433. doi:https://doi.org/10.1105/tpc.105.031831.

Dean, J.D., Goodwin, P.H. and Hsiang, T. (2002). Comparison of relative RT-PCR and northern blot analyses to measure expression of β-1,3-glucanase inNicotiana benthamiana infected withColltotrichum destructivum. *Plant Molecular Biology Reporter*, 20(4), pp.347–356. doi:https://doi.org/10.1007/bf02772122.

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), pp.1935–1940. doi:https://doi.org/10.1016/j.cub.2004.10.028.

Douady, C.J. (2003). Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. *Molecular Biology and Evolution*, 20(2), pp.248–254. doi:https://doi.org/10.1093/molbev/msg042.

Egea-Cortines, M., Saedler, H. and Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is

involved in the control of floral architecture in Antirrhinum majus. *The EMBO Journal*, 18(19), pp.5370–5379. doi:https://doi.org/10.1093/emboj/18.19.5370.

Eichorn, S.E. and Evert, R.F. (2013). *Raven biology of plants*. 8th ed. New York: W.H. Freeman.

ENDRESS, P. and DOYLE, J. (2007). Floral phyllotaxis in basal angiosperms: development and evolution. *Current Opinion in Plant Biology*, 10(1), pp.52–57. doi:https://doi.org/10.1016/j.pbi.2006.11.007.

Endress, P.K. (2001). Origins of flower morphology. *Journal of Experimental Zoology*, 291(2), pp.105–115. doi:https://doi.org/10.1002/jez.1063.

Endress, P.K. and Doyle, J.A. (2009). Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany*, 96(1), pp.22–66. doi:https://doi.org/10.3732/ajb.0800047.

Fang, Y., Qin, X., Liao, Q., Du, R., Luo, X., Zhou, Q., Li, Z., Chen, H., Jin, W., Yuan Ya-Ning, Sun, P., Zhang, R., Zhang, J., Wang, L., Cheng, S., Yang, X., Yan, Y., Zhang, X., Zhang, Z. and Bai, S.-N. (2022). The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nature Plants*, 8(9), pp.1024–1037. doi:https://doi.org/10.1038/s41477-022-01222-x.

Feng, Y.-Y., Du, H., Huang, K.-Y., Ran, J.-H. and Wang, X.-Q. (2024). Reciprocal expression of MADS-box genes and DNA methylation reconfiguration initiate bisexual cones in spruce. *Communications Biology*, 7(1). doi:https://doi.org/10.1038/s42003-024-05786-6.

Finch-Savage, W.E. and Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. *New Phytologist*, 171(3), pp.501–523. doi:https://doi.org/10.1111/j.1469-8137.2006.01787.x.

Frankema, E. (2014). Africa and the Green Revolution A Global Historical Perspective. *NJAS - Wageningen Journal of Life Sciences*, 70-71, pp.17–24. doi:https://doi.org/10.1016/j.njas.2014.01.003.

Galletta, B.J. and Rusan, N.M. (2015). A Yeast Two-Hybrid approach for probing protein-protein interactions at the centrosome. *Methods Cell Bio*, 129, pp.251–277. doi:https://doi.org/10.1016/bs.mcb.2015.03.012.

Ganger, M.T., Girouard, J.A., Smith, H.M., Bahny, B.A. and Ewing, S.J. (2015). Antheridiogen and abscisic acid affect conversion and ANI1 expression in Ceratopteris richardii gametophytes. *Botany*, 93(2), pp.109–116. doi:https://doi.org/10.1139/cjb-2014-0202.

Gao, L., Lee, J.S., Sariel Hübner, Hulke, B.S., Qu, Y. and Rieseberg, L.H. (2019). Genetic and phenotypic analyses indicate that resistance to flooding stress is uncoupled from performance in cultivated sunflower. *New Phytologist*, 223(3), pp.1657–1670. doi:https://doi.org/10.1111/nph.15894.

Ghannam, M.G. and Varacallo, M. (2018). *Biochemistry, Polymerase Chain Reaction (PCR)*. [online] Nih.gov. Available at: https://www.ncbi.nlm.nih.gov/books/NBK535453/.

Glass, F. and Takenaka, M. (2018). The Yeast Three-Hybrid System for Protein Interactions. *Methods in molecular biology*, pp.195–205. doi:https://doi.org/10.1007/978-1-4939-7871-7_12.

Gleave, A.P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology*, 20(6), pp.1203–1207. doi:https://doi.org/10.1007/bf00028910.

Gligorijević, V., Renfrew, P.D., Kosciolek, T., Leman, J.K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B.C., Fisk, I.M., Vlamakis, H., Xavier, R.J., Knight, R., Cho,

K. and Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. *Nature Communications*, 12(1). doi:https://doi.org/10.1038/s41467-021-23303-9.

Goto , K., Kyozuka , J. and Bowman , J. (2001). Turning floral organs into leaves, leaves into floral organs. *Curr Opin Genet Dev.*, 11(4), pp.449–456. doi:https://doi.org/10.1016/s0959-437x(00)00216-1..

Gramzow, L. and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biology*, 11(6), p.214. doi:https://doi.org/10.1186/gb-2010-11-6-214.

Gramzow, L. and Theißen, G. (2013). Phylogenomics of MADS-Box Genes in Plants — Two Opposing Life Styles in One Gene Family. *Biology*, 2(3), pp.1150–1164. doi:https://doi.org/10.3390/biology2031150.

Gramzow, L. and Theißen, G. (2015). Phylogenomics reveals surprising sets of essential and dispensable clades of MIKCc-group MADS-box genes in flowering plants. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324(4), pp.353–362. doi:https://doi.org/10.1002/jez.b.22598.

Gramzow, L., Barker, E., Schulz, C., Ambrose, B., Ashton, N., Theißen, G. and Litt, A. (2012). Selaginella Genome Analysis – Entering the 'Homoplasy Heaven' of the MADS World. *Frontiers in Plant Science*, 3. doi:https://doi.org/10.3389/fpls.2012.00214.

Gramzow, L., Ritz, M.S. and Theißen, G. (2010). On the origin of MADS-domain transcription factors. *Trends in Genetics*, 26(4), pp.149–153. doi:https://doi.org/10.1016/j.tig.2010.01.004.

Harrison, C.J. and Morris, J.L. (2017). The origin and early evolution of vascular plant shoots and leaves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1739), p.20160496. doi:https://doi.org/10.1098/rstb.2016.0496.

Hasebe, M., Wen, C.-K., Kato, M. and Banks, J.A. (1998). Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proceedings of the National Academy of Sciences*, 95(11), pp.6222–6227. doi:https://doi.org/10.1073/pnas.95.11.6222.

Hatfield, J.L. and Prueger, J.H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10(Part A), pp.4–10. doi:https://doi.org/10.1016/j.wace.2015.08.001.

He, C., Si, C., Teixeira da Silva, J.A., Li, M. and Duan, J. (2019). Genome-wide identification and classification of MIKC-type MADS-box genes in Streptophyte lineages and expression analyses to reveal their role in seed germination of orchid. *BMC Plant Biology*, 19(1). doi:https://doi.org/10.1186/s12870-019-1836-5.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S. and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation . *Plant Molecular Biology*, 42(6), pp.819–832. doi:https://doi.org/10.1023/a:1006496308160.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T. and Theißen, G. (2002). Two Ancient Classes of MIKC-type MADS-box Genes are Present in the Moss Physcomitrella patens. *Molecular Biology and Evolution*, 19(6), pp.801–814. doi:https://doi.org/10.1093/oxfordjournals.molbev.a004137.

Hileman, L.C., Sundstrom, J.F., Litt, A., Chen, M., Shumba, T. and Irish, V.F. (2006). Molecular and Phylogenetic Analyses of the MADS-Box Gene Family in Tomato. *Molecular Biology and Evolution*, 23(11), pp.2245–2258. doi:https://doi.org/10.1093/molbev/msl095.

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, [online] 409(6819), pp.525–529. doi:https://doi.org/10.1038/35054083.

Hsu, W., Yeh, T., Huang, K., Li, J., Chen, H. and Yang, C. (2013). *AGAMOUS-LIKE13*, a putative ancestor for the E functional genes, specifies male and female gametophyte morphogenesis. *The Plant Journal*, 77(1), pp.1–15. doi:https://doi.org/10.1111/tpj.12363.

Huang, B., Kong, L., Wang, C., Ju, F., Zhang, Q., Zhu, J., Gong, T., Zhang, H., Yu, C., Zheng, W.-M. and Bu, D. (2023). Protein Structure Prediction: Challenges, Advances, and the Shift of Research Paradigms. *Genomics, Proteomics & Bioinformatics*. doi:https://doi.org/10.1016/j.gpb.2022.11.014.

Huang, X., Wang, W., Gong, T., Wickell, D., Kuo, L.-Y., Zhang, X., Wen, J., Kim, H., Lu, F., Zhao, H., Chen, S., Li, H., Wu, W., Yu, C., Chen, S., Fan, W., Chen, S., Bao, X., Li, L. and Zhang, D. (2022). The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nature Plants*, [online] 8(5), pp.500–512. doi:https://doi.org/10.1038/s41477-022-01146-6.

Immink, R., Busscher, M., Angenent, G.C., Gadella, T.W.J. and Ferrario, S. (2002). Analysis of MADS box protein–protein interactions in living plant cells. *Proceedings of the National Academy of Sciences of the United States of America*, 99(4), pp.2416–2421. doi:https://doi.org/10.1073/pnas.042677699.

Immink, R.G.H., Kaufmann, K. and Angenent, G.C. (2010). The ‘ABC’ of MADS domain protein behaviour and interactions. *Seminars in Cell & Developmental Biology*, 21(1), pp.87–93. doi:https://doi.org/10.1016/j.semcdb.2009.10.004.

Irish, V. (2017). The ABC model of floral development. *Current Biology*, [online] 27(17), pp.R887–R890. doi:https://doi.org/10.1016/j.cub.2017.03.045.

Ismail Elhassnaoui, Wahba, S., Sinafekesh Girma Wolde, Bahaj, M. and Aniss Moumen (2023). The past, current, and future of the Africa Green Revolution: The case study of Kenya, Morocco, and Nigeria. *World Water Policy*. doi:https://doi.org/10.1002/wwp2.12135.

Jalali, M., Zaborowska, J. and Jalali, M. (2017). *Chapter 1 - The Polymerase Chain Reaction: PCR, qPCR, and RT-PCR*. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/B9780128030776000011.

Jetha, K., Theißen, G. and Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Research*, [online] 42(17), pp.10927–10942. doi:https://doi.org/10.1093/nar/gku755.

Jiang, W., Deng, F., Babla, M., Chen, C., Yang, D., Tong, T., Qin, Y., Chen, G., Marchant, B., Soltis, P., Douglas Edward Soltis, Zeng, F. and Chen, Z.-H. (2024). Efficient gene editing of a model fern species through gametophyte-based transformation. *PLANT PHYSIOLOGY*. doi:https://doi.org/10.1093/plphys/kiae473.

John, D.A. and Babu, G.R. (2021). Lessons From the Aftermaths of Green Revolution on Food System and Health. *Frontiers in Sustainable Food Systems*, [online] 5(1). doi:https://doi.org/10.3389/fsufs.2021.644559.

Jones, V.A.S. and Dolan, L. (2012). The evolution of root hairs and rhizoids. *Annals of Botany*, [online] 110(2), pp.205–212. doi:https://doi.org/10.1093/aob/mcs136.

Kapoor, B., Kumar, P., Verma, V., Irfan, M., Sharma, R. and Bhargava, B. (2022). How plants conquered land: evolution of terrestrial adaptation. *Journal of Evolutionary Biology*, 35(5). doi:https://doi.org/10.1111/jeb.14062.

Kaufmann, K., Melzer, R. and Theißen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene*, 347(2), pp.183–198. doi:https://doi.org/10.1016/j.gene.2004.12.014.

Kellogg, E.A. (2001). Evolutionary History of the Grasses. *Plant Physiology*, [online] 125(3), pp.1198–1205. doi:https://doi.org/10.1104/pp.125.3.1198.

Kerridge, E. (2013). *The Agricultural Revolution*. Abingdon: Routledge.

Khehra, N., Padda, I.S. and Swift, C.J. (2023). *Polymerase Chain Reaction (PCR)*. [online] PubMed. Available at: https://www.ncbi.nlm.nih.gov/books/NBK589663/.

Kinosian, S.P. and Wolf, P.G. (2022). The biology of C. richardii as a tool to understand plant evolution. *eLife*, [online] 11, p.e75019. doi:https://doi.org/10.7554/eLife.75019.

Kiselev, K.V., Aleynova, O.A., Ogneva, Z.V., Suprun, A.R. and Dubrovina, A.S. (2021). 35S promoter-driven transgenes are variably expressed in different organs of Arabidopsis thaliana and in response to abiotic stress. *Molecular Biology Reports*, 48(3), pp.2235–2241. doi:https://doi.org/10.1007/s11033-021-06235-x.

Knie, N., Fischer, S., Grewe, F., Polsakiewicz, M. and Knoop, V. (2015). Horsetails are the sister group to all other monilophytes and Marattiales are sister to leptosporangiate ferns. *Molecular Phylogenetics and Evolution*, 90, pp.140–149. doi:https://doi.org/10.1016/j.ympev.2015.05.008.

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). doi:https://doi.org/10.1093/bioinformatics/btz305.

Krizek, B.A. and Fletcher, J.C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics*, [online] 6(9), pp.688–698. doi:https://doi.org/10.1038/nrg1675.

Krizek, B.A. and Meyerowitz, E.M. (1996). The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development*, [online] 122(1), pp.11–22. doi:https://doi.org/10.1242/dev.122.1.11.

KROGAN, N.T. and ASHTON, N.W. (2000). Ancestry of plant MADS-box genes revealed by bryophyte (Physcomitrella patens) homologues. *New Phytologist*, 147(3), pp.505–517. doi:https://doi.org/10.1046/j.1469-8137.2000.00728.x.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), pp.1547–1549. doi:https://doi.org/10.1093/molbev/msy096.

Kwantes, M., Liebsch, D. and Verelst, W. (2011). How MIKC* MADS-Box Genes Originated and Evidence for Their Conserved Function Throughout the Evolution of Vascular Plant Gametophytes. *Molecular Biology and Evolution*, [online] 29(1), pp.293–302. doi:https://doi.org/10.1093/molbev/msr200.

Lai, X., Daher, H., Galien, A., Hugouvieux, V. and Zubieta, C. (2019). Structural Basis for Plant MADS Transcription Factor Oligomerization. *Computational and Structural Biotechnology Journal*, [online] 17, pp.946–953. doi:https://doi.org/10.1016/j.csbj.2019.06.014.

Lai, X., Vega-Léon, R., Hugouvieux, V., Romain Blanc-Mathieu, van, Lucas, J., Silva, C.S., Jourdain, A., Muino, J.M., Nanao, M.H., Immink, R., Kaufmann, K., François Parcy, Cezary Smaczniak and Zubieta, C. (2021). The intervening domain is required for DNA-binding and functional identity of plant MADS transcription factors. *Nature communications*, 12(1). doi:https://doi.org/10.1038/s41467-021-24978-w.

Leslie, A.B. and Bonacorsi, N.K. (2022). Understanding the appearance of heterospory and derived plant reproductive strategies in the Devonian. *Paleobiology*, 48(3), pp.496–512. doi:https://doi.org/10.1017/pab.2021.44.

Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A., Koppers, N., Kuo, L.-Y., Li, Z., Simenc, M., Small, I., Wafula, E., Angarita, S., Barker, M.S., Bräutigam, A., dePamphilis, C., Gould, S., Hosmani, P.S. and Huang, Y.-M. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, 4(7), pp.460–472. doi:https://doi.org/10.1038/s41477-018-0188-8.

Linkies, A., Graeber, K., Knight, C. and Leubner-Metzger, G. (2010). The evolution of seeds. *New Phytologist*, 186(4), pp.817–831. doi:https://doi.org/10.1111/j.1469-8137.2010.03249.x.

Litt, A. and Irish, V.F. (2003). Duplication and Diversification in the APETALA1/FRUITFULL Floral Homeotic Gene Lineage: Implications for the Evolution of Floral Development. *Genetics*, 165(2), pp.821–833. doi:https://doi.org/10.1093/genetics/165.2.821.

Liu, S., Wang, H. and Qin, F. (2023). Genetic dissection of drought resistance for trait improvement in crops. *The Crop Journal*, [online] 11(4), pp.975–985. doi:https://doi.org/10.1016/j.cj.2023.05.002.

Livak, K.J. and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, 25(4), pp.402–408.

Lorek, T.W. (2022). The Green Revolution in Latin America. *Oxford Research Encyclopedia of Latin American History*. doi:https://doi.org/10.1093/acrefore/9780199366439.013.1085.

Lotthammer, J.M., Ginell, G.M., Griffith, D., Emenecker, R.J. and Holehouse, A.S. (2024). Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nature Methods*, [online] 21(3), pp.465–476. doi:https://doi.org/10.1038/s41592-023-02159-5.

Louche, A., Salcedo, S.P. and Bigot, S. (2017). Protein-Protein Interactions: Pull-Down Assays. *Methods in Molecular Biology (Clifton, N.J.)*, [online] 1615, pp.247–255. doi:https://doi.org/10.1007/978-1-4939-7033-9_20.

Lovisetto, A., Guzzo, F., Tadiello, A., Toffali, K., Favretto, A. and Casadoro, G. (2011). Molecular Analyses of MADS-Box Genes Trace Back to Gymnosperms the Invention of Fleshy Fruits. *Molecular Biology and Evolution*, 29(1), pp.409–419. doi:https://doi.org/10.1093/molbev/msr244.

Ma, H., Yanofsky, M.F. and Meyerowitz, E.M. (1991). AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes & Development*, 5(3), pp.484–495. doi:https://doi.org/10.1101/gad.5.3.484.

Madeira, F., Nandana Madhusoodanan, Lee, J., Eusebi, A., Ania Niewielska, Tivey, A.R.N., Lopez, R. and Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic acids research*. doi:https://doi.org/10.1093/nar/gkae241.

Malcomber, S.T. and Kellogg, E.A. (2005). SEPALLATA gene diversification: brave new whorls. *Trends in Plant Science*, 10(9), pp.427–435. doi:https://doi.org/10.1016/j.tplants.2005.07.008.

Mansoury, M., Hamed, M., Karmustaji, R., Al Hannan, F. and Safrany, S.T. (2021). The edge effect: A global problem. The trouble with culturing cells in 96-well plates. *Biochemistry and Biophysics Reports*, 26, p.100987. doi:https://doi.org/10.1016/j.bbrep.2021.100987.

Marchant, D.B., Chen, G., Cai, S., Chen, F., Schafran, P., Jenkins, J., Shu, S., Plott, C., Webber, J., Lovell, J.T., He, G., Sandor, L., Williams, M., Rajasekar, S., Healey, A., Barry, K., Zhang, Y., Sessa, E., Dhakal, R.R. and Wolf, P.G. (2022). Dynamic genome evolution in a model fern. *Nature Plants*, [online] 8(9), pp.1038–1051. doi:https://doi.org/10.1038/s41477-022-01226-7.

Marchant, D.B., Sessa, E.B., Wolf, P.G., Heo, K., Barbazuk, W.B., Soltis, P.S. and Soltis, D.E. (2019). The C-Fern (Ceratopteris richardii) genome: insights into plant genome evolution with the first partial homosporous fern genome assembly. *Scientific Reports*, 9(1). doi:https://doi.org/10.1038/s41598-019-53968-8.

Margaret Anne Pelayo, Yamaguchi, N. and Ito, T. (2021). One factor, many systems: the floral homeotic protein AGAMOUS and its epigenetic regulatory mechanisms. *Current Opinion in Plant Biology*, 61, pp.102009–102009. doi:https://doi.org/10.1016/j.pbi.2021.102009.

Martienssen, R. and Moazed, D. (2015). RNAi and Heterochromatin Assembly. *Cold Spring Harbor Perspectives in Biology*, 7(8), p.a019323. doi:https://doi.org/10.1101/cshperspect.a019323.

Mason, J.M. and Arndt, K.M. (2004). Coiled Coil Domains: Stability, Specificity, and Biological Implications. *ChemInform*, 35(15). doi:https://doi.org/10.1002/chin.200415270.

McCormack, J.E., Huang, H. and Knowles, L.L. (2009). Maximum Likelihood Estimates of Species Trees: How Accuracy of Phylogenetic Inference Depends upon the Divergence History and Sampling Design. *Systematic Biology*, 58(5), pp.501–508. doi:https://doi.org/10.1093/sysbio/syp045.

McLoughlin, S., Carpenter, R.J., Jordan, G.J. and Hill, R.S. (2008). Seed ferns survived the end-Cretaceous mass extinction in Tasmania. *American Journal of Botany*, 95(4), pp.465–471. doi:https://doi.org/10.3732/ajb.95.4.465.

Medard, N. and Yanofsky, Ma.F. (2001). Function and evolution of the plant MADS-box gene family. *Nature Reviews*, 2, pp.186–195.

Meng, D., Cao, Y., Chen, T., Abdullah, M., Jin, Q., Fan, H., Lin, Y. and Cai, Y. (2019). Evolution and functional divergence of MADS-box genes in Pyrus. *Scientific Reports*, [online] 9(1). doi:https://doi.org/10.1038/s41598-018-37897-6.

Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H. and Ferrin, T.E. (2023). UCSF ChimeraX: Tools for Structure Building and Analysis. *Protein Science: A Publication of the Protein Society*, [online] 32(11), p.e4792. doi:https://doi.org/10.1002/pro.4792.

Messenguy, F. and Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development . *Gene*, 16(316), pp.1–21. doi:https://doi.org/10.1016/s0378-1119(03)00747-9.

Miki, D. and Shimamoto, K. (2004). Simple RNAi Vectors for Stable and Transient Suppression of Gene Function in Rice. *Plant and Cell Physiology*, 45(4), pp.490–495. doi:https://doi.org/10.1093/pcp/pch048.

Miller, K., Eggenberger, A.L., Lee, K., Liu, F., Kang, M., Drent, M., Ruba, A., Kirscht, T., Wang, K. and Jiang, S. (2021). An improved biolistic delivery and analysis method for evaluation of DNA and CRISPR-Cas delivery efficacy in plant tissue. *Scientific Reports*, [online] 11(1), p.7695. doi:https://doi.org/10.1038/s41598-021-86549-9.

Milner, G.R. and Boldsen, J.L. (2023). Population trends and the transition to agriculture: Global processes as seen from North America. *Proceedings of the National Academy of Sciences*, 120(4). doi:https://doi.org/10.1073/pnas.2209478119.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, [online] 19(6), pp.1–4. doi:https://doi.org/10.1038/s41592-022-01488-1.

Mizukami, Y. and Ma, H. (1995). Separation of AG function in floral meristem determinacy from that in reproductive organ identity by expressing antisense AG RNA. *Plant Molecular Biology*, 28(5), pp.767–784. doi:https://doi.org/10.1007/bf00042064.

Mizukami, Y., Huang, H., Tudor, M., Hu, Y. and Ma, H. (1996). Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. *The Plant Cell*, 8(5), pp.831–845. doi:https://doi.org/10.1105/tpc.8.5.831.

Möckli, N., Deplazes, A., Hassa, P.O., Zhang, Z., Peter, M., Hottiger, M.O., Igor Stagljar and Auerbach, D. (2007). Yeast Split-Ubiquitin-Based Cytosolic Screening System to Detect Interactions Between Transcriptionally Active Proteins. *BioTechniques*, 42(6), pp.725–730. doi:https://doi.org/10.2144/000112455.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H. and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 115(10), pp.E2274–E2283. doi:https://doi.org/10.1073/pnas.1719588115.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C. and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biology*, 9(Suppl 1), p.S4. doi:https://doi.org/10.1186/gb-2008-9-s1-s4.

Munster, T., Pahnke, J., Di Rosa, A., Kim, J.T., Martin, W., Saedler, H. and Theissen, G. (1997). Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proceedings of the*

National Academy of Sciences, 94(6), pp.2415–2420. doi:https://doi.org/10.1073/pnas.94.6.2415.

Murai, K. (2013). Homeotic Genes and the ABCDE Model for Floral Organ Formation in Wheat. *Plants*, 2(3), pp.379–395. doi:https://doi.org/10.3390/plants2030379.

Nam, J., dePamphilis , C.W., Ma, H. and Nei , M. (2003). Antiquity and Evolution of the MADS-Box Gene Family Controlling Flower Development in Plants. *Molecular Biology and Evolution*, 20(9), pp.1435–1447. doi:https://doi.org/10.1093/molbev/msg152.

Nam, J., Kim, J., Lee, S., An, G., Ma, H. and Nei, M. (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proceedings of the National Academy of Sciences*, [online] 101(7), pp.1910–1915. doi:https://doi.org/10.1073/pnas.0308430100.

Ng, M. and Yanofsky, M.F. (2001). Function and evolution of the plant MADS-box gene family. *Nature Reviews Genetics*, 2(3), pp.186–195. doi:https://doi.org/10.1038/35056041.

Ning, K., Han, Y., Chen, Z., Luo, C., Wang, S., Zhang, W., Li, L., Zhang, X., Su, F. and Wang, Q. (2019). Genome-wide analysis of MADS-box family genes during flower development in lettuce. *Plant Cell and Environment*, 42(6), pp.1868–1881. doi:https://doi.org/10.1111/pce.13523.

Ohad, N., Shichrur, K. and Yalovsky, S. (2007). The Analysis of Protein-Protein Interactions in Plants by Bimolecular Fluorescence Complementation. *PLANT PHYSIOLOGY*, 145(4), pp.1090–1099. doi:https://doi.org/10.1104/pp.107.107284.

Olsson, O. and Paik, C. (2016). Long-run cultural divergence: Evidence from the Neolithic Revolution. *Journal of Development Economics*, 122, pp.197–213. doi:https://doi.org/10.1016/j.jdeveco.2016.05.003.

Parenicova, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C. and Colombo, L. (2003). Molecular and Phylogenetic Analyses of the Complete MADS-Box Transcription Factor Family in Arabidopsis: New Openings to the MADS World. *THE PLANT CELL ONLINE*, 15(7), pp.1538–1551. doi:https://doi.org/10.1105/tpc.011544.

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*, [online] 405(6783), pp.200–203. doi:https://doi.org/10.1038/35012103.

Pellegrini, L., Tan, S. and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. *Nature*, [online] 376(6540), pp.490–498. doi:https://doi.org/10.1038/376490a0.

Pereira, A.M. and Coimbra, S. (2019). Advances in plant reproduction: from gametes to seeds. *Journal of Experimental Botany*, 70(11), pp.2933–2936. doi:https://doi.org/10.1093/jxb/erz227.

phytozome-next.jgi.doe.gov. (n.d.). *Phytozome v13*. [online] Available at: https://phytozome-next.jgi.doe.gov.

Pingali, P. (2012). Green Revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences*, [online] 109(31), pp.12302–12308. doi:https://doi.org/10.1073/pnas.0912953109.

Plackett, A.R., Conway, S.J., Hewett Hazelton, K.D., Rabbinowitsch, E.H., Langdale, J.A. and Di Stilio, V.S. (2018). LEAFY maintains apical stem cell activity during shoot development in the fern Ceratopteris richardii. *eLife*, [online] 7, p.e39625. doi:https://doi.org/10.7554/eLife.39625.

Plackett, A.R.G., Huang, L., Sanders, H.L. and Langdale, J.A. (2014). High-Efficiency Stable Transformation of the Model Fern Species Ceratopteris richardii via

Microparticle Bombardment . *Plant Physiology*, 165(1), pp.3–14. doi:https://doi.org/10.1104/pp.113.231357.

Plackett, A.R.G., Rabbinowitsch, E.H. and Langdale, J.A. (2015). Protocol: genetic transformation of the fern Ceratopteris richardii through microparticle bombardment. *Plant Methods*, 11(1). doi:https://doi.org/10.1186/s13007-015-0080-8.

Pratt, A.I., Knoblauch, J. and Kunz, H.-H. (2020). An updated pGREEN-based vector suite for cost-effective cloning in plant molecular biology. *PubMed*, [online] 2020. doi:https://doi.org/10.17912/micropub.biology.000317.

Pray, L.A. (2008). The Biotechnology Revolution: PCR and the Use of Reverse Transcriptase to Clone Expressed Genes. *Nature Education*, 1(1).

Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E., Hart, D., Theißen, G., Silva, C.S., Parcy, F., Dumas, R., Nanao, M. and Zubieta, C. (2014). Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in Arabidopsis . *The Plant Cell*, 26(9), pp.3603–3615. doi:https://doi.org/10.1105/tpc.114.127910.

Qaim, M. and Kouser, S. (2013). Genetically Modified Crops and Food Security. *PLoS ONE*, [online] 8(6). doi:https://doi.org/10.1371/journal.pone.0064879.

Qiu, Y., Li, Z. and Köhler, C. (2024). Ancestral duplication of MADS-box genes in land plants empowered the functional divergence between sporophytes and gametophytes. *New Phytologist*. doi:https://doi.org/10.1111/nph.20065.

Qiu, Y., Li, Z., Walther, D. and Köhler, C. (2023). Updated Phylogeny and Protein Structure Predictions Revise the Hypothesis on the Origin of MADS-box Transcription Factors in Land Plants. *Molecular biology and evolution*, 40(9). doi:https://doi.org/10.1093/molbev/msad194.

Raza, A., Razzaq, A., Mehmood, S.S., Zou, X., Zhang, X., Lv, Y. and Xu, J. (2019). Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. *Plants*, [online] 8(2). doi:https://doi.org/10.3390/plants8020034.

Reece-Hoyes, J.S. and Walhout, A.J.M. (2018). Gateway Recombinational Cloning. *Cold Spring Harbor Protocols*, 2018(1), p.pdb.top094912. doi:https://doi.org/10.1101/pdb.top094912.

Reeves, P.A. and Richards, C.M. (2007). Distinguishing Terminal Monophyletic Groups from Reticulate Taxa: Performance of Phenetic, Tree-Based, and Network Procedures. *Systematic Biology*, 56(2), pp.302–320. doi:https://doi.org/10.1080/10635150701324225.

Rensing, S.A. (2018). Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology*, 42, pp.49–54. doi:https://doi.org/10.1016/j.pbi.2018.02.006.

Rensing, S.A., Goffinet, B., Meyberg, R., Wu, S.-Z. and Bezanilla, M. (2020). The Moss Physcomitrium (Physcomitrella) patens: A Model Organism for Non-Seed Plants. *The Plant Cell*, 32(5), pp.1361–1376. doi:https://doi.org/10.1105/tpc.19.00828.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F. ., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S. . and Yamaguchi, K. (2007). The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science*, 319(5859), pp.64–69. doi:https://doi.org/10.1126/science.1150646.

Riechmann, J.L. and Meyerowitz, E.M. (1997a). Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent

of their DNA-binding specificity. *Molecular Biology of the Cell*, 8(7), pp.1243–1259. doi:https://doi.org/10.1091/mbc.8.7.1243.

Riechmann, J.L. and Meyerowitz, E.M. (1997b). MADS domain proteins in plant development. *PubMed*, 378(10), pp.1079–101.

Rijpkema, A.S., Vandenbussche, M., Koes, R., Heijmans, K. and Gerats, T. (2010). Variations on a theme: Changes in the floral ABCs in angiosperms. *Seminars in Cell & Developmental Biology*, [online] 21(1), pp.100–107. doi:https://doi.org/10.1016/j.semcdb.2009.11.002.

Rutherford, G., Tanurdzic, M., Hasebe, M. and Banks, J. (2004). A systemic gene silencing method suitable for high throughput, reverse genetic analyses of gene function in fern gametophytes. *BMC Plant Biology*, 4(1), p.6. doi:https://doi.org/10.1186/1471-2229-4-6.

Satu Ruokolainen, Yan Peng Ng, Albert, V.A., Elomaa, P. and Teeri, T.H. (2010). Large scale interaction analysis predicts that the Gerbera hybrida floral E function is provided both by general and specialized proteins. *BMC Plant Biology*, 10(1). doi:https://doi.org/10.1186/1471-2229-10-129.

Schmittgen, T.D. and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3(6), pp.1101–1108. doi:https://doi.org/10.1038/nprot.2008.73.

Schoe, D.J., Johnson, M.T.J. and Wright, S.I. (2019). The ecology, evolution, and genetics of plant reproductive systems. *New Phytologist*, 224(3), pp.99–1004.

Selwyn, M., Pino, J. and Espelta, J.M. (2023). Disentangling the importance of intrinsic and extrinsic seed dispersal factors in forest restoration success: a global review. *Restoration Ecology*. doi:https://doi.org/10.1111/rec.13868.

Shen, G., Yang, C.-H., Shen, C.-Y. and Huang, K.-S. (2019). Origination and selection of ABCDE and AGL6 subfamily MADS-box genes in gymnosperms and angiosperms. *Biological Research*, 52(1). doi:https://doi.org/10.1186/s40659-019-0233-8.

Shivhare, D., Musialak-Lange, M., Julca, I., Gluza, P. and Mutwil, M. (2021). Removing auto-activators from yeast-two-hybrid assays by conditional negative selection. *Scientific Reports*, 11(1). doi:https://doi.org/10.1038/s41598-021-84608-9.

Shizuka Koshimizu, Rumiko Kofuji, Sasaki-Sekimoto, Y., Masahide Kikkawa, Mie Shimojima, Ohta, H., Shuji Shigenobu, Kabeya, Y., Yuji Hiwatashi, Yosuke Tamada, Murata, T. and Hasebe, M. (2018). Physcomitrella MADS-box genes regulate water supply and sperm movement for fertilization. *Nature Plants*, 4(1), pp.36–45. doi:https://doi.org/10.1038/s41477-017-0082-9.

Simossis, V.A. and Heringa, J. (2003). The PRALINE online server: optimising progressive multiple alignment on the web. *Computational Biology and Chemistry*, 27(4-5), pp.511–519. doi:https://doi.org/10.1016/j.compbiolchem.2003.09.002.

Simossis, V.A. and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Research*, 33(Web Server), pp.W289–W294. doi:https://doi.org/10.1093/nar/gki390.

Singer, S.D., Krogan, N.T. and Ashton, N.W. (2007). Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. *Plant Cell Reports*, 26(8), pp.1155–1169. doi:https://doi.org/10.1007/s00299-007-0312-0.

Smaczniak, C., Immink, R.G.H., Angenent, G.C. and Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development*, 139(17), pp.3081–3098. doi:https://doi.org/10.1242/dev.074674.

Smidkova, M., Hola, M. and Angelis, K.J. (2010). Efficient biolistic transformation of the moss Physcomitrella patens. *Biologia plantarum*, 54(4), pp.777–780. doi:https://doi.org/10.1007/s10535-010-0141-9.

Soltis, D.E., Chanderbali, A.S., Kim, S., Buzgo, M. and Soltis, P.S. (2007). The ABC Model and its Applicability to Basal Angiosperms. *Annals of Botany*, 100(2), pp.155–163. doi:https://doi.org/10.1093/aob/mcm117.

Sourcebioscience.com. (2024). *Source BioScience Genomics*. [online] Available at: https://genomics.sourcebioscience.com/ [Accessed 26 Sep. 2024].

Specht, C.D. and Bartlett, M.E. (2009). Flower Evolution: The Origin and Subsequent Diversification of the Angiosperm Flower. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), pp.217–243. doi:https://doi.org/10.1146/annurev.ecolsys.110308.120203.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), pp.1312–1313. doi:https://doi.org/10.1093/bioinformatics/btu033.

Stuessy, T. and Weiss-Schneeweiss, H. (2019). What drives polyploidization in plants? *New Phytologist*, 223(4), pp.1690–1692. doi:https://doi.org/10.1111/nph.15929.

Svensson, M.E. and Engstrom, P. (2002). Closely related MADS-box genes in club moss (Lycopodium) show broad expression patterns and are structurally similar to, but phylogenetically distinct from, typical seed plant MADS-box genes. *New Phytologist*, 154(2), pp.439–450. doi:https://doi.org/10.1046/j.1469-8137.2002.00392.x.

Tan, L. and Yammani, R.R. (2022). Co-Immunoprecipitation-Blotting: Analysis of Protein-Protein Interactions. *Methods in Molecular Biology (Clifton, N.J.)*, [online] 2413, pp.145–154. doi:https://doi.org/10.1007/978-1-0716-1896-7_15.

Tan, S. and Richmond, T.J. (1998). Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. *Nature*, 391(6668), pp.660–666. doi:https://doi.org/10.1038/35563.

Tanabe, Y., Hasebe, M., Sekimoto, H., Nishiyama, T., Kitani, M., Henschel, K., Munster, T., Theissen, G., Nozaki, H. and Ito, M. (2005). Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proceedings of the National Academy of Sciences*, 102(7), pp.2436–2441. doi:https://doi.org/10.1073/pnas.0409860102.

Tanabe, Y., Uchida, M., Hasebe, M. and Ito, M. (2002). Characterization of the Selaginella remotifolia MADS-box gene. *Journal of Plant Research*, 116(1), pp.69–73. doi:https://doi.org/10.1007/s10265-002-0071-5.

Thangavel, G. and Nayar, S. (2018). A Survey of MIKC Type MADS-Box Genes in Non-seed Plants: Algae, Bryophytes, Lycophytes and Ferns. *Frontiers in Plant Science*, 9. doi:https://doi.org/10.3389/fpls.2018.00510.

Theißen, G. and Gramzow, L. (2016). Structure and Evolution of Plant MADS Domain Transcription Factors. *Plant Transcription Factors*, pp.127–138. doi:https://doi.org/10.1016/b978-0-12-800854-6.00008-7.

Theißen, G. and Saedler, H. (2001). Floral quartets. *Nature*, [online] 409(6819), pp.469–471. doi:https://doi.org/10.1038/35054172.

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Münster, T., Winter, K.-U. and Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Molecular Biology*, 42(1), pp.115–149. doi:https://doi.org/10.1023/a:1006332105728.

Theißen, G., Kim, J.T. and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution*, 43(5), pp.484–516. doi:https://doi.org/10.1007/bf02337521.

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*, [online] 143(18), pp.3259–3271. doi:https://doi.org/10.1242/dev.134080.

United Nations (2019). *Plants, the 'core Basis for Life on Earth', under Increasing threat, Warns UN Food Agency*. [online] UN News. Available at: https://news.un.org/en/story/2019/12/1052591.

Vajda, V. and McLoughlin, S. (2004). Fungal Proliferation at the Cretaceous-Tertiary Boundary. *Science*, 303(5663), pp.1489–1489. doi:https://doi.org/10.1126/science.1093807.

Vekemans, D., Viaene, T., Caris, P. and Geuten, K. (2011). Transference of function shapes organ identity in the dove tree inflorescence. *New Phytologist*, 193(1), pp.216–228. doi:https://doi.org/10.1111/j.1469-8137.2011.03915.x.

Vesty, E.F., Saidi, Y., Moody, L.A., Holloway, D., Whitbread, A., Needs, S., Choudhary, A., Burns, B., McLeod, D., Bradshaw, S.J., Bae, H., King, B.C., Bassel, G.W., Simonsen, H.T. and Coates, J.C. (2016). The decision to germinate is regulated by divergent molecular networks in spores and seeds. *New Phytologist*, 211(3), pp.952–966. doi:https://doi.org/10.1111/nph.14018.

Vinicius Costa Galvao and Schmid, M. (2014). Regulation of Flowering by Endogenous Signals. *Advances in Botanical Research*, pp.63–102. doi:https://doi.org/10.1016/b978-0-12-417162-6.00003-1.

Wang, B., Xiao, Y., Yan, M., Fan, W., Zhu, Y., Li, W. and Li, T. (2024). Gene Duplication and Functional Diversification of MADS-Box Genes in Malus × domestica following WGD: Implications for Fruit Type and Floral Organ Evolution. *International*

Journal of Molecular Sciences, 25(16), p.8962. doi:https://doi.org/10.3390/ijms25168962.

Wang, P., Liao, H., Zhang, W., Yu, X., Zhang, R., Shan, H., Duan, X., Yao, X. and Kong, H. (2015a). Flexibility in the structure of spiral flowers and its underlying mechanisms. *Nature Plants*, 2(1). doi:https://doi.org/10.1038/nplants.2015.188.

Wang, X. and Bai, S.-N. (2019). Key innovations in transition from homospory to heterospory. *Plant Signaling & Behavior*, 14(6), p.1596010. doi:https://doi.org/10.1080/15592324.2019.1596010.

Wang, X., Wang, J., Jin, D., Guo, H., Lee, T.-H., Liu, T. and Paterson, Andrew H. (2015b). Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events. *Molecular Plant*, 8(6), pp.885–898. doi:https://doi.org/10.1016/j.molp.2015.04.004.

Wang, Y.-Q., Melzer, R. and Theißen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm Gnetum gnemon provide a clue to the evolutionary origin of 'floral quartets'. *The Plant Journal*, 64(2), pp.177–190. doi:https://doi.org/10.1111/j.1365-313x.2010.04325.x.

West, A.G., Sharrocks, A.D., Causier, B.E. and Davies, B. (1998). DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors. *Nucleic Acids Research*, 26(23), pp.5277–5287. doi:https://doi.org/10.1093/nar/26.23.5277.

Whisstock, J.C. and Lesk, A.M. (2003). Prediction of protein function from protein sequence and structure. *Quarterly Reviews of Biophysics*, 36(3), pp.307–340. doi:https://doi.org/10.1017/s0033583503003901.

Wilson, J.P., White, J.D., Montañez, I.P., DiMichele, W.A., McElwain, J.C., Poulsen, C.J. and Hren, M.T. (2020). Carboniferous plant physiology breaks the mold. *The New Phytologist*, [online] 227(3), pp.667–679. doi:https://doi.org/10.1111/nph.16460.

Withers, K.A., Falls, K., Youngstrom, C.E., Nguyen, T., DeWald, A., Yarvis, R.M., Simons, G.P., Flanagan, R., Bui, L.T., Irish, E.E. and Cheng, C.-L. (2023). A Ceratopteris EXCESS MICROSPOROCYTES1 suppresses reproductive transition in the fern vegetative leaves. *Plant Science*, 335, pp.111812–111812. doi:https://doi.org/10.1016/j.plantsci.2023.111812.

Wolberger, C. (1998). Combinatorial transcription factors. *Current Opinion in Genetics & Development*, 8(5), pp.552–559. doi:https://doi.org/10.1016/s0959-437x(98)80010-5.

Wu, W., Chen, F., Jing, D., Liu, Z. and Ma, L. (2011). Isolation and Characterization of an AGAMOUS-Like Gene from Magnolia wufengensis (Magnoliaceae). *Plant Molecular Biology Reporter*, 30(3), pp.690–698. doi:https://doi.org/10.1007/s11105-011-0385-3.

Wu, Y., You, H.-L. and Li, X.-Q. (2017). Dinosaur-associated Poaceae epidermis and phytoliths from the Early Cretaceous of China. *National Science Review*, 5(5), pp.721–727. doi:https://doi.org/10.1093/nsr/nwx145.

Xfam.org. (2018). *Pfam*. [online] Available at: https://pfam.xfam.org/.

Xiang, D.-L. and Li, G.-S. (2024). Control of leaf development in the water fern Ceratopteris richardii by the auxin efflux transporter CrPINMa in the CRISPR/Cas9 analysis. *BMC Plant Biology*, 24(1). doi:https://doi.org/10.1186/s12870-024-05009-4.

Yang, Y., Fanning, L. and Jack, T. (2003). The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. *The Plant Journal*, 33(1), pp.47–59. doi:https://doi.org/10.1046/j.0960-7412.2003.01473.x.

Yang, Z. and Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, 13(5), pp.303–314. doi:https://doi.org/10.1038/nrg3186.

Youngstrom, C.E., Withers, K.A., Irish, E.E. and Cheng, C.-L. (2022). Vascular function of the T3/modern clade WUSCHEL-Related HOMEOBOX transcription factor genes predate apical meristem-maintenance function. *BMC Plant Biology*, 22(1). doi:https://doi.org/10.1186/s12870-022-03590-0.

Yuan, L. and Perry, S.E. (2011). *Plant Transcription Factors*. *Methods in molecular biology*, New York: Springer Science+Business Media, pp.3–9. doi:https://doi.org/10.1007/978-1-61779-154-3.

Zahn, L.M., Kong, H., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., dePamphilis, C.W. and Ma, H. (2005). The Evolution of theSEPALLATASubfamily of MADS-Box Genes. *Genetics*, 169(4), pp.2209–2223. doi:https://doi.org/10.1534/genetics.104.037770.

Zhang, L., Zhao, J., Feng, C., Liu, M., Wang, J. and Hu, Y. (2017). Genome-wide identification, characterization of the MADS-box gene family in Chinese jujube and their involvement in flower development. *Scientific Reports*, 7(1). doi:https://doi.org/10.1038/s41598-017-01159-8.

Zhang, X., Pan, L., Guo, W., Li, Y. and Wang, W. (2022). A convergent mechanism of sex determination in dioecious plants: Distinct sex-determining genes display converged regulation on floral B-class genes. *Frontiers in Plant Science*, 13. doi:https://doi.org/10.3389/fpls.2022.953445.

Zhang, X., Wang, Q., Yang, S., Lin, S., Bao, M., Bendahmane, M., Wu, Q., Wang, C. and Fu, X. (2018). Identification and Characterization of the MADS-Box Genes and Their Contribution to Flower Organ in Carnation (Dianthus caryophyllus L.). *Genes*, [online] 9(4), p.193. doi:https://doi.org/10.3390/genes9040193.

Zhang, Y., Chen, M., Siemiatkowska, B., Toleco, M.R., Jing, Y., Strotmann, V., Zhang, J., Stahl, Y. and Fernie, A.R. (2020). A Highly Efficient Agrobacterium-Mediated Method for Transient Gene Expression and Functional Studies in Multipe Plant Species. *Plant Communications*, p.100028. doi:https://doi.org/10.1016/j.xplc.2020.100028.

Zhao, P.-X., Zhang, J., Chen, S.-Y., Wu, J., Xia, J.-Q., Sun, L.-Q., Ma, S.-S. and Xiang, C.-B. (2021). Arabidopsis MADS-box factor AGL16 is a negative regulator of plant response to salt stress by downregulating salt-responsive genes. *The New Phytologist*, [online] 232(6), pp.2418–2439. doi:https://doi.org/10.1111/nph.17760.

Zhong, S., Tian, Y., Guan, J., Zhang, Q., Chen, C. and Luo, P. (2024). Expression and function analyses of the MIKCC-type MADS-box genes in Akebia trifoliata (Lardizabalaceae) flower development. *Acta Physiologiae Plantarum*, 46(5). doi:https://doi.org/10.1007/s11738-024-03683-3.

Zobell, O., Faigl, W., Saedler, H. and Munster, T. (2010). MIKC* MADS-Box Proteins: Conserved Regulators of the Gametophytic Generation of Land Plants. *Molecular Biology and Evolution*, 27(5), pp.1201–1211. doi:https://doi.org/10.1093/molbev/msq005.

APPENDIX 1

1. Recipes for growth media and buffers

1.1. Growth media

C. fern medium liquid (1L)

- 100 ml macronutrients (10x solution)
- 10 ml Fe-EDTA (100x solution)
- 5 ml micronutrients (200x solution)
- Made up to 1L with dH_2O
- Adjust to pH 6.0 using KOH

C. fern medium agar (1L)

- 100 ml macronutrients (10x solution)
- 10 ml Fe-EDTA (100x solution)
- 5 ml micronutrients (200x solution)
- 10 g plant culture agar
- Made up to 1L with dH_2O
- Adjust to pH 6.0 using KOH

Macronutrients 10x (1L)

- 1.25 g NH_4NO_3
- 5 g KH_2PO_4
- 1.2 g $MgSO_4$-$7H_2O$
- 0.26 g $CaCl_2$-$2H_2O$
- Made up to 1L with dH_2O

Micronutrients 200x (1L)

- 0.05 g $MnSO_4$-H_2O
- 0.074 g $CuSO_4$-$5H_2O$
- 0.104 g $ZnSO_4$-$7H_2O$
- 0.372 g H_3BO_3
- 0.0074 g $(NH_4)Mo_6O_{24}$-$4H_2O$
- Made up to 1L with dH_2O

Chelated Iron 100x (1L)

- 2.78 g $FeSO_4$-$7H_2O$
- 3.73 g Disodium EDTA-$2H_2O$
- Made up to 1L with dH_2O

LB liquid (1L)

- 10 g bacto-tryptone
- 5 g bacto-yeast extract
- 10 g NaCl
- Made up to 1L with dH_2O

LB agar (1L)

- 10 g bacto-tryptone
- 5 g bacto-yeast extract
- 10 g NaCl
- 10 g bacto-agar
- Made up to 1L with dH_2O

1x MS 2% sucrose medium (1L)

- 0.5 g plant MES hydrate
- 4.16 g Murashige-Skoog
- 20 g sucrose
- 7 g agar
- Made up to 1L with dH_2O
- Adjust to pH 5.8 using KOH

Yeast drop-out media (500 mL)

For Yeast drop-out media the appropriate Takara Yeast two-hybrid media pouch (SD-LT, SD-ALT, SD-HLT, SD-AHLT) was diluted with 0.5 L dH_2O and autoclaved at 121 °C for 20 minutes.

Yeast drop-out YPDA medium (1L)

- 20 g bacto-peptone
- 10 g yeast extract
- 20 g sucrose/dextrose
- 20 g agar
- 40 mg adenine
- Made up to 1L with dH_2O

1.2. Buffers

1M Tris-HCl pH 9.0 (400 mL)

- 48.46g Tris-HCl (MW: 121.4)
- Adjust pH to 9.0 with 1 M HCl
- Make up to 400 mL with dH_2O

2% CTAB extraction buffer (400 mL)

- 40 ml 1M Tris-HCl pH 9.0
- 16 ml 500 mM EDTA
- 8 g CTAB (Hexadecyltrimethylammonium bromide)
- 32.72 g NaCl
- Made up to 400 mL with dH_2O

0.5M EDTA pH 8.0 (400 ml final volume):

- 58.448 g EDTA (MW: 292.24)
- Adjust to pH 8.0 using KOH pellets
- Made up to 400 mL with dH_2O

TAE 50x stock solution (1L)

- 242 g Tris base (MW = 121)
- 57.1 mL glacial acetic acid
- 100 mL 0.5 M EDTA (pH 8.0)
- Made up to 1 L with dH_2O

APPENDIX 2

2. Primer table

Table 1: Primers used to amplify *CMADS1* and it's potential interactors, designed using Phytozome genome database (https://phytozome-next.jgi.doe.gov).

Gene	**Sequence 5' to 3'**
CerMADS1 Forward	GATTCAATCCGTTAAGGAGATTCTG
CerMADS1 Reverse	GGTATAATAGAGCCTCACCGTC
CerMADS1 new primers Forward	GGAAGCTATGGTGAGGACG
CerMADS1 new primers Reverse	CTATGCACAAGGTTTTAGTGGTATAATAG
CMADS1 Forward	CATCTCCTGCAGCTCTAAGAC
CMADS1 Reverse	CAGCTACAACCTCAGGTCAAG
CMADS2 Forward	CGATGGTGAGGAGGAAGATC
CMADS2 Reverse	CTATTCCATACAAACTGGTGAGTG
CMADS2 new Forward	GAGGCTGCGAGCCTC
CMADS2 new Reverse	GAGCTTTGGATATCTGTTCTATTCC
CMADS3 Forward	CATTCTGCTCATACACCAGC
CMADS3 Reverse	GAATTTCCTTAGCGCCTTGC
CMADS4 Forward	CACAGCACGAACATCCTAC
CMADS4 Reverse	CCATCCACAGCCTTAATCATG
CMADS4 new primers Forward	GTTTTGCAGATCCGCTCG
CMADS4 new primers Reverse	CATCCACAGCCTTAATCATGATTC
CMADS6 Forward	GTTGGTGAGTCGATTCCATG
CMADS6 Reverse	CCATTAGTTCAGTCTCAAGTCGAG
CRM1 Forward	CTCATACACCAGCAGTGGAC

CRM1 Reverse	GAATTTCCTTAGCGCCTTGC
CRM2 Forward	CTGATATCGCGACGGTAGAG
CRM2 Reverse	CTGAATTCCTGTCTTATTCCATGC
CrMADS_A Forward	GCAATGGTGAGGAGGAAGATC
CrMADS_A Reverse	TTATGCAGACGAACAAAAATCAACTAG
CrMADS_AG Forward	GAGAGAGAGAGAGAATCGATGG
CrMADS_AG Reverse	CATTTAAGAGAGATTTCACAGTGATAGTAG
CrMADS_AG new primers Forward	GAGAGAGAATCGATGGGAAGG
CrMADS_AG new primers Reverse	GAAAAACGTGGATCGTTGAGAG
CrMADS_AI Forward	GTGAGATGATTTCGAGTGCTG
CrMADS_AI Reverse	CAACAGAGACATGACCATCAAAG
CrMADS_AI new primers Forward	GACTACGTGAGATGATTTCGAGTG
CrMADS_AI new primers Reverse	CATCAAAGCTTGCGTTCCTG
CrMADS_AJ Forward	GTTTTGCAGACCTGCTCG
CrMADS_AJ Reverse	CAGTAACCATCCACAACGTAAATC
CrMADS_E Forward	CGTAGATGGCATTAAACACACG
CrMADS_E Reverse	CTTACATTCGAGGACATCCAAG
CrMADS_G Forward	GATATGGGCCGTTGCAAG
CrMADS_G Reverse	GTATCAATACTTGCATATGTTGTCC
CrMADS_H Forward	CAGGCAGGAACACAAATACAG
CrMADS_H Reverse	GAGAAAGTGGCTGACATTCTG
CrMADS_I Forward	CGATGGTGAGGAGGAAAATC
CrMADS_I Reverse	CTACTCAGACGCAGGTGATTC
CrMADS_M Forward	ATGGCAATGCCATGTTCAG

CrMADS_M Reverse	TCATTGTGTGAAGCCAAGC
CrMADS_S Forward	ATGGCCTTGGATAAAAGCG
CrMADS_S Reverse	TCACCCTAGCTTCAGCG
CrMADS_W Forward	CATTCTCCATTGGAACTGCG
CrMADS_W Reverse	CTTCGACCGCATCCTGG
V4_38160 Forward	GATTGAGAACCCGATGACG
V4_38160 Reverse	CCTGTCTTATTCCATGCAAAG
V4_51781 Forward	CTTCGCTCCGCATCATTTC
V4_51781 Reverse	CTCGTAAGAGGAACGTCTGC

Table 2: Primers used to amplify both Type I and Type II *Ceratopteris richardii* MADS-box genes for genome validation.

Primer name	**Sequence 5' to 3'**
M13 Forward	CACCTCTACTTTCGTCTCTCTC
M13 Reverse	CAATGTGAACAGCTTCCGTC
M14 Forward	CCGGTGAAAGTGTCAGAAG
M14 Reverse	CTAGCTTTGCTTTAGAACGCTC
M14 Internal Forward	CTCTCGGTATTGAGGGACTATC
M14 Internal Reverse	GCATATCACTCACATGGCAC
M15 Forward	CCTTGGAGAATTACTTTGAGCG
M15 Reverse	CAGATTGAACACCGCTTACC
M16 Forward	CTGGCAAGAATCAGCATTGAC
M16 Reverse	CAATGTCCTCGATTCAATGAAACC

CrMADS_C Forward	GATATGCAACTACCGAGGC
CrMADS_C Reverse	GATTCATTCAACGAGAGTCTGAG
CrMADS_D Forward	GAAATGGGTCGCGCTAAAATC
CrMADS_D Reverse	TTATTTGTGGGTTTGCTCGTG
CrMADS_F Forward	GTGGCAATTTGCAGAAATGG
CrMADS_F Reverse	CCTAGCCGAATTCCTCTGTAAC
CrMADS_J Forward	ATGGGGAGGGTCAAGCTTG
CrMADS_J Reverse	TTAGGGGCAATCGAAAATCAGAAG
CrMADS_L Forward	ATGGGGAGGGTCAAGC
CrMADS_L Reverse	TTATCGCCGTCGTGAAACAG
CrMADS_N Forward	ATGGCAATGCCATGTTCAG
CrMADS_N Reverse	CTATAACATGATAAAACGTGAAACATAAGTC
CrMADS_O Forward	ATGGGGAGGGTGAAGC
CrMADS_O Reverse	TTATCTACTCCGCATAGACTTTTC
CrMADS_Q Forward	GAGAATGGCAGCCACAG
CrMADS_Q Reverse	TCATGCATCAATGCCATCTG
CrMADS_Q new primers Forward	GCGAGTGAGAATGGCAG
CrMADS_Q new primers Reverse	GTGATTGTCGTGTCTTCATGC
CrMADS_R Forward	CATGGGGAAAAGCAAGATTTTG
CrMADS_R Reverse	GCTATTACTGAATGTTGAGAACACC
NRCDS_015287 Forward	CTCGTCGGGTCTTATGGTG
NRCDS_015287 Reverse	CTATCTTCGCACTGTCGAAAAG
V4_89546 Forward	GGCGATCTATGCTGATAGATG

V4_89546 Reverse	CATGCTTAATCTGATGATGTGG

Table 3: Primers used for Yeast-II-Hybrid:

Primer name	**Sequence 5’ to 3’**
CMADS1-F-SfiI	GGCCATGGAGGCCCATCTCCTG
CMADS1-R-XmaI	CCCGGGCAGCTACAACCTCAGG
CMADS1-F-SfiI (no start codon)	CTGCTTGGCCTCCATGGCC
CMADS1-R-XmaI (no stop codon)	GACCTGAGGTTGCCCGGG
CMADS1-F-XmaI-no-start	CCCGGGAAGCAGCTTCCG
pGBKT7 T7 promoter forward sequencing primer	GTAATACGACTCACTATAGGGCGA
pGBKT7 T7 terminator reverse sequencing primer	CCCCAAGGGGTTATGCTAG
pGADT7 T7 promoter forward sequencing primer	TAATACGACTCACTATAGGGCG
pGADT7 reverse sequencing primer	CAGTATCTACGATTCATCTGC
CMADS1-F-MIK-no-start-XmaI	CCCGGGAATACTTTCCTCTC
CMADS1-R-MIK-XmaI	CCCGGGTATAATTGTTGCC

Table 4: primers used for qPCR and RNAi Gateway cloning

Gene	**Primer sequence**
Actin B Forward	GAGAGAGGCTACTCTTTCACAACC
Actin B Reverse	AGGAAGTTCGTAACTCTTCTCCAA
E02 (putative TATA box) Forward	ATGAGCCAGAGCTTTTCCCC
E02 (putative TATA box) Reverse	TTCGTCTCTGACCTTTGCCC
M08 (putative Ubiquitin9) Forward	TCAATCTGCTCCCTGCTCAC
M08 (putative Ubiquitin9) Reverse	GTCCAGTTCCTTGCTGTTGC
CMADS1-qPCR-F	GAACCAGATACCACTGGATCTAGCTG
CMADS1-qPCR-R	GTTCTCCTCAAATAGCCTTGCCTC
CrMADS_S-qPCR-F	CAGGCGATATGAGGATCTCCAATCTC
CrMADS_S-qPCR-R	GCAGCCGCATTTTCGCCC
CrMADS_H-qPCR-F	CCGCATAGTGCACATCGCG
CrMADS_H-qPCR-R	GCAGTTCTTGAAATGATAAGCCTGTTG
CrMADS_M-qPCR-F	GATCATCTTCTCCAGCACAGGAAAAC
CrMADS_M-qPCR-R	GCTCAGCCTCTTGCTTCCAG
CMADS1-F-SfiI	GGCCATGGAGGCCCATCTCCTG
CMADS1-R-XmaI	CCCGGGCAGCTACAACCTCAGG
CMADS1-F-SfiI (no start codon)	CTGCTTGGCCTCCATGGCC
CMADS1-R-XmaI (no stop codon)	GACCTGAGGTTGCCCGGG
CMADS1-F-SfiI-no-start	GGCCATGGAGGCCAAGCAGC
CMADS1-F-XmaI-no-start	CCCGGGAAGCAGCTTCCG

Table 5: RNA interference primers:

	CMADS1 Primer 5' to 3'	AttB1/AttB2 5' to 3'	Construct 5' to 3'
Forward primer	GAATGGGGCTCTGGTACA	GGGGACAAGTTTGTACAAAAAAGCAGGCT	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAATGGGGCTCTGGTACA
Reverse primer	CTGCGGATCGGAATTGTTC	GGGGACCACTTTGTACAAGAAAGCTGGGT	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGCGGATCGGAATTGTTC

3. *Ceratopteris* RNA extracted for use in cDNA synthesis and the associated assays

Table 6: *Ceratopteris richardii* RNA samples extracted from different developmental stages of tissues.

RNA Sample Name	**number of individuals**	**nano drop concentrations**
CrWT RNA 1 27/11/20	20	65.2 ng/µl

CrWT RNA 2 27/11/20	10	45.3 ng/µl
CrWT RNA 3 27/11/20	5	19.3 ng/µl
RNA 01/12/20 10 day hermaphrodite (n) 4	20	89.2 ng/µl
RNA 01/12/20 10 day hermaphrodite (n) 5	10	24.7 ng/µl
RNA 01/12/20 10 day hermaphrodite (n) 6	5	18.3 ng/µl
CrWT (n) 5 day male and hermaphrodite RNA 150	150	11.2 ng/µl
1 CrWT (n) 7 day hermaphrodite 50 RNA 8/12/20	50	18.1 ng/µl
2 CrWT (n) 7 day hermaphrodite 100 RNA 8/12/20	100	16.5 ng/µl
3 CrWT (n) 7 day hermaphrodite 150 RNA 8/12/20	150	33.1 ng/µl
4 CrWT (n) 7 day male 50 RNA	50	10.7 ng/µl
4 CrWT (2n) RNA 9 days 5 18/12/20 RNA	5	68.3 ng/µl
5 CrWT (2n) RNA 9 day 18/12/20 10 RNA	10	99.8 ng/µl
6 CrWT (2n) RNA 9 day 18/12/20 15 RNA	15	205.8 ng/µl
1 CrWT (n) RNA hermaphrodite 14 day 30 RNA	30	63.0 ng/µl
2 CrWT (n) RNA hermaphrodite 14 day 3018/12/20 RNA	30	60.7 ng/µl
3 CrWT (n) RNA hermaphrodite 14 day 30 18/12/20 RNA	30	68.4 ng/µl
8 CrWT (n) 14d hermaphrodite 30 15/01/21 RNA	30	22.9 ng/µl
1 CrWT (n) 7d hermaphrodite 13/01/21 RNA 300	300	31.9 ng/µl
2 CrWT (n) 7d hermaphrodite 13/01/21 RNA 300	300	61.4 ng/µl
1 CrWT (n) 12d male 13/01/21 RNA 25	25	9.7 ng/µl
2 CrWT (n) 12d male 13/01/21 RNA 50	50	16.5 ng/µl
3 CrWT (n) 12d male 13/01/21 RNA 100	100	12.9 ng/µl

4 CrWT (n) 12d male 13/01/21 RNA 25	25	24.5 ng/µl
5 CrWT (n) 12d male 13/01/21 RNA 50	50	36.5 ng/µl
6 CrWT (n) 12d male 13/01/21 RNA 100	100	43.7 ng/µl
1 CrWT 2n mature fronds 15/01/21 RNA 10	10	58.7 ng/µl
2 CrWT 2n mature fronds 15/01/21 RNA 10	10	20.4 ng/µl
3 CrWT 2n mature fronds 15/01/21 RNA 10	10	22.9 ng/µl
4 CrWT 2n mature fronds 15/01/21 RNA 10	10	36.1 ng/µl
5 CrWT 2n expanding fronds 15/01/21 RNA 10	10	25.8 ng/µl
6 CrWT 2n expanding fronds 15/01/21 RNA 10	10	43.3 ng/µl
7 CrWT 2n expanding fronds 15/01/21 RNA 10	10	50.4 ng/µl
2 CrWT (n) 5d male and hermaphrodite RNA 500 individuals 01/02/21	250	65.0 ng/µl
Combined with above for a 500 individual sample	250	-
Combined with 2 CrWT (n) 5d male and hermaphrodite 250 individuals 09/02/21	250	
1 CrWT (n) 7d male 500 individual RNA 01/02/21	250	21.1 ng/µl
Combined with above for a 500 individual sample	250	-
1 CrWT (n) 10d male 500 individual RNA 01/02/21	250	54.0 ng/µl
Combined with above for a 500 individual sample	250	-
2 CrWT (n) 5d male and hermaphrodite RNA 500 individuals 11/02/21	250	97.1 ng/µl
3 CrWT (n) 5d male and germaphrodite RNA 500 individuals 11/02/21	250	
Combined with above for a 500 individual sample	250	74.8 ng/µl

2 CrWT (n) 10 day male 500 individuals 30/04/21	250	129.7 ng/µl
Combined with the above for a 500 individual sample	250	
2 CrWT (n) 7 day male 500 individuals 30/04/21	250	119.5 ng/µl
Combined with above for a 500 individual sample	250	
3 CrWT (n) 7 day hermaphrodite 500 individuals 30/04/21	250	120.2 ng/µl
Combined with the above for a 500 individual sample	250	
1 CrWT (2n) 130 day reproductive fiddle head 04/02/22	3	609.7 ng/µl
2 CrWT (2n) 130 day reproductive fiddle head 04/02/22	3	368.8 ng/µl
3 CrWT (2n) 130 day reproductive fiddle head 04/02/22	4	375.4 ng/µl
1 CrWT (2n) 130 day reproductive expanding frond 04/02/22	3	1464.4 ng/µl
2 CrWT (2n) 130 day reproductive expanding frond 10/02/22	3	1426.4 ng/µl
3 CrWT (2n) 130 day reproductive expanding frond/02/22	3	1217.4 ng/µl
1 CrWT (2n) 130 day reproductive mature frond 10/02/22	1	401.3 ng/µl

2 CrWT (2n) 130 day reproductive mature frond 10/02/22	1	353.7 ng/µl
3 CrWT (2n) 130 day reproductive mature frond 10/02/22	1	346.6 ng/µl
1 CrWT (2n) Vegetative expanding fronds 63 day 26/07/2022	3	1020.6 ng/µl
2 CrWT (2n) Vegetative expanding fronds 63 day 26/07/2022	3	1026.5 ng/µl
3 CrWT (2n) Vegetative expanding fronds 63 day 26/07/2022	3	1996.7 ng/µl
1 CrWT (2n) Vegetative mature fronds 63 day 26/07/2022	3	203.9 ng/µl
2 CrWT (2n) Vegetative mature fronds 63 day 26/07/2022	3	950.3 ng/µl
3 CrWT (2n) Vegetative mature fronds 63 day 26/07/2022	3	1211.0 ng/µl
4 CrWT (2n) Vegetative mature fronds 63 day 26/07/2022	3	342.6 ng/µl
1 CrWT (2n) Vegetative fiddle head 63 day 26/07/2022	2	287.2 ng/µl
2 CrWT (2n) Vegetative fiddle head 63 day 13/01/2023	4	915.3 ng/µl
3 CrWT (2n) Vegetative fiddle head 63 day 13/01/2023	4	55.6 ng/µl

4 CrWT (2n) Vegetative fiddle head 63 day 13/01/2023	4	524.4 ng/µl

APPENDIX 3

4. Overexpression of *CMADS1* in *Ceratopteris*

4.1. Restriction cloning troubleshooting of tagged *35S::CMADS1*-myc

The initial cloning of CMADS1:myc:pJET showed three potential PCR errors with one causing an amino acid change. A further PCR was performed and the subsequent mini-preps were sequenced. These had no mismatch base pairs, although one read was slightly short. Internal primers showed the correct sequence was reliable throughout. Restriction sites KpnI and HindIII were then cloned into pJ*Et al*ong with the myc tagged CMADS1 insert. This construct was then transformed into the pART7 vector, this vector is used to donate the 35S promotor region to be used in overexpression of the chosen gene insertion, as well as an OCS terminator region.

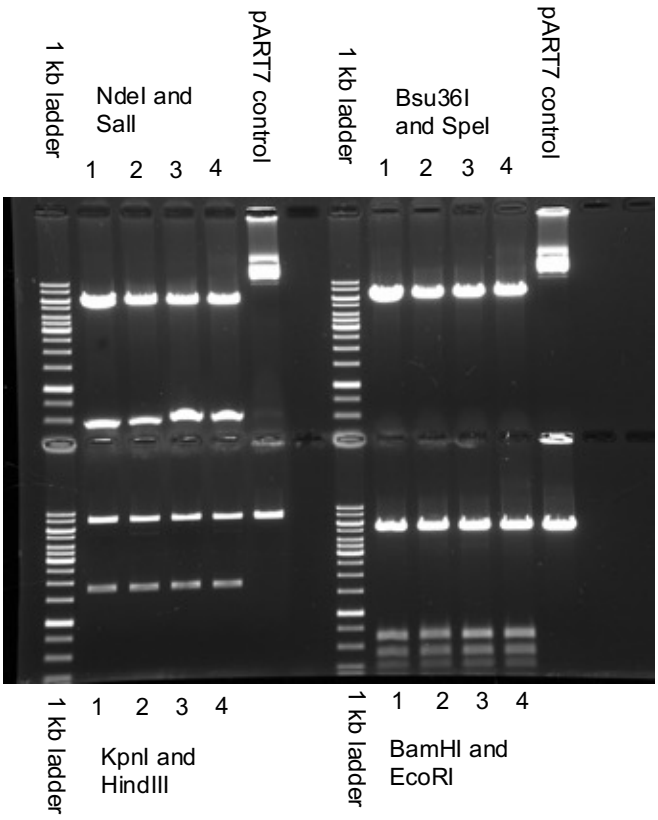


Figure 1: Double diagnostic digests of CMADS1-myc in the intermediate vector pART7, 4 transformations were performed and ran on 0.8% agarose electrophoresis gel at a constant 90V for 60 minutes. The expected bands when digested with NdeI and SalI when an insertion was present was approximately 6 kb and 450 bp. The expected bands when digested with Bsu361 and SpeI when an insertion was present was approximately 6 kb, 800 bp, and 700 bp. The expected bands when digested with KpnI and HinIII when an insertion was present was approximately 5 kb and 1.5 kb. The expected bands when digested with BamHI and EcoRI when an insertion was present was approximately 5 kb, 600 bp, 400 bp, and 250 bp. An unsuccessful insertion with any of the restriction enzyme pairs would give only a 6 kb band.

Diagnostic digests showed all four mini preps had successful insertions of myc tagged CMADS1. Mini-preps were then digested with NotI at 37C overnight. As only one restriction site was used for the insertion into pBOMBER a check for orientation was performed.

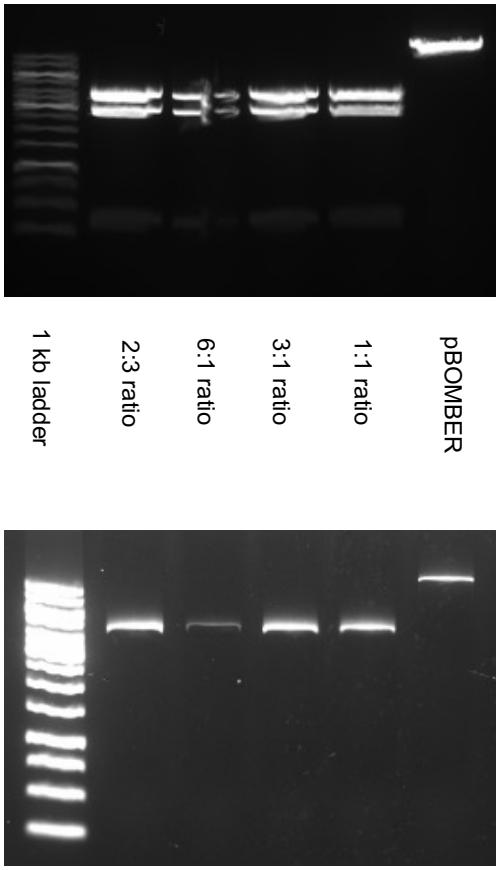
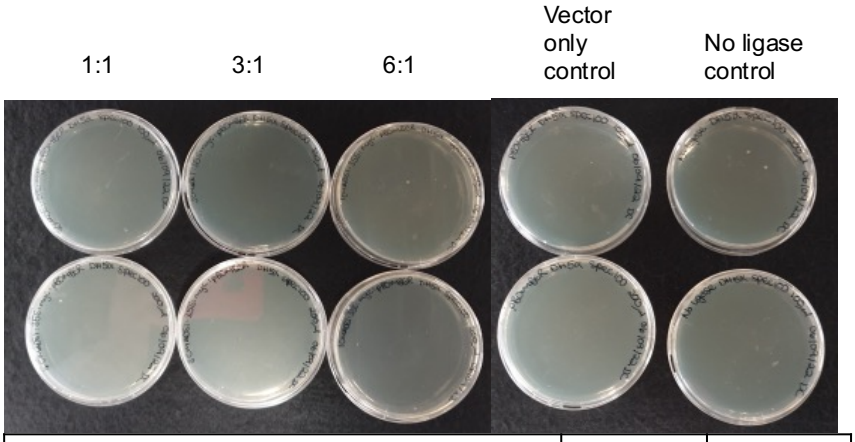


Figure 2: the top gel shows 4 digests of 35S::CMADS1-myc in pART7 using NotI restriction enzyme, with the final lane shows a pBOMBER sample being digested with NotI. The bottom gel image shows the purified DNA from the digestion gel. The initial ligation calculation suggested a ratio of 2:3 vector:insert, however as the purified gel showed similar intensities for both the vector DNA and the insert DNA so a 1:1 ratio was also performed. As all insertions were successful previously extra ratios of 3:1 and 6:1 were also performed. A 90-minute ligation at 37C was performed for the 3:1 and 1:1 ratios, a 2 hour ligation at 37C was performed for 6:1 and 2:3 ratios.

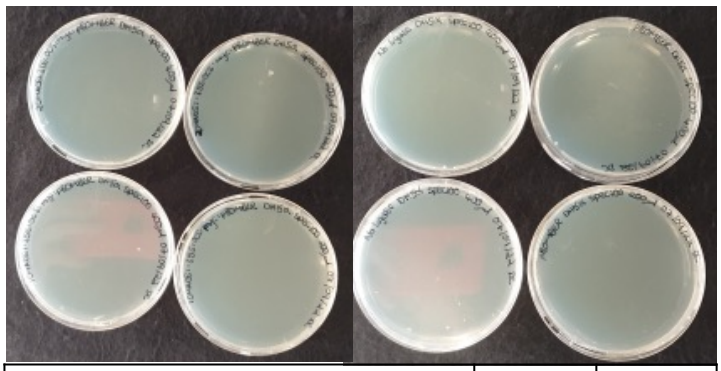
The above ligations were transformed in DH5α following the lab heat shock protocol described in chapter 2. After 24 hours incubating at 37C no growth could be seen. After 48 hours no growth could be seen on the 2:3 ratio plate. The 1:1, 3:1, and 6:1 plates had limited growth as is shown in figure 3.



	100 µl	200 µl
6:1 *35S::CMADS1*-myc in pBOMBER	1	1
3:1 *35S::CMADS1*-myc in pBOMBER	0	0
1:1 *35S::CMADS1*-myc in pBOMBER	0	1
pBOMBER vector only control	0	0
No ligase control	0	2

Figure 3: colony plates of *35S::CMADS1*-myc in pBOMBER cloned into DHα, the table shows colony counts of all ratio plates along with control plates.

Mini-prep 1 and 2 were used for another transformation. For the ligation stage a 2-hour incubation at 37C was performed and the ratios of insert:vector were 6:1 and 2:3. The ratio 2:3 was used due to the purified DNA gel (figure 2) showing a lower concentration of vector compared to the insert. The volume plated was doubled as the previous transformation did show an increase in colonies between 100 µl and 200 µl.



	200 µl	400 µl
1 pBOMBER::35S::CMADS1-myc	0	1
2 pBOMBER::35S::CMADS1-myc	0	1
pBOMBER vector only control	2	1
No ligase control	0	0

Figure 4: colony plates of *35S::CMADS1*-myc in pBOMBER at the revised insert:vector ratios with a higher volume of transformation culture plated.

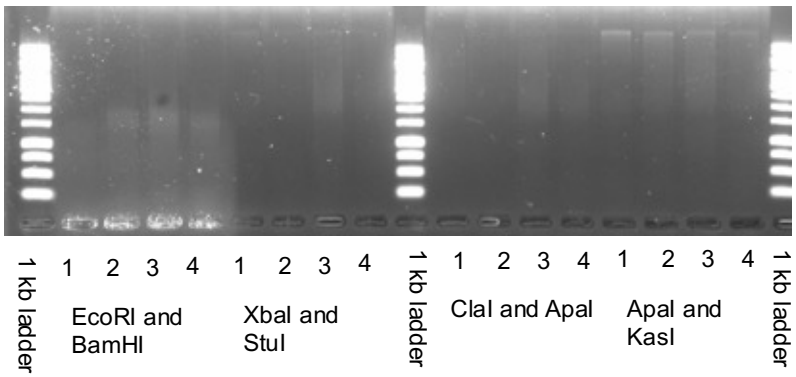


Figure 5: A diagnostic digest electrophoresis gel of *35S::CMADS1*-myc in pBOMBER The following expected bands: EcoRI and BamHI: Insertion: ~13.6 kb, ~600 bp, ~400 bp, and ~250 bp. XbaI and StuI: Insertion: ~12.8 kb, and ~2.1 kb. ClaI and ApaI: ~ 13.3 kb, and ~ 1.7 kb. ApaI and KasI: ~ 9 kb, and ~5.8 kb. Without an insertion the expected band would be approximately 11.5 kb.

4.2. Restriction cloning troubleshooting of tagged *35S::CMADS1*

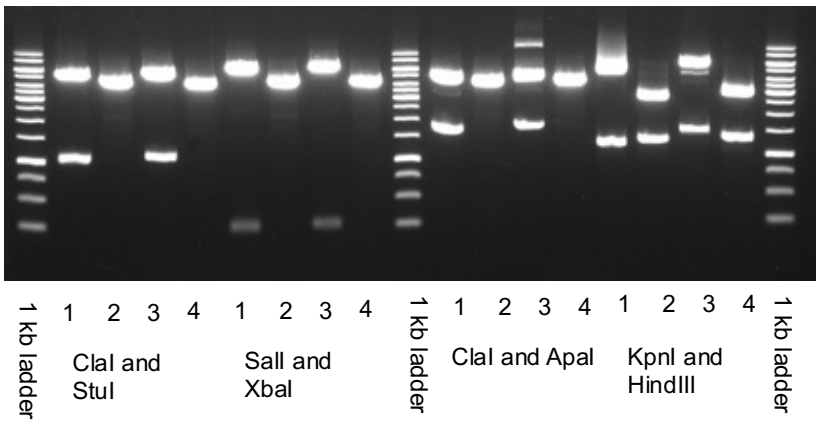


Figure 6: A diagnostic digest electrophoresis gel of 4 pairs of double digestions with *CMADS1* in pART7. The expected bands for each digestion being the following: ClaI and StuI: successful insertion: approximately 5.2 kb, and 1 kb. SalI and XbaI successful insertion: approximately 6 kb, and 240 bp. ClaI and ApaI successful insertion approximately 4.6 kb, and 1.6 kb. KpnI and HindIII successful insertion approximately 4.5 kb, and ~1.3 kb. An unsuccessful insertion with any digest would give an expected band of approximately 6 kb.

The diagnostic digests of CMADS1 in pART7 mini-preps suggest samples 1 and 3 had successful insertions whilst 2 and 4 were unsuccessful.

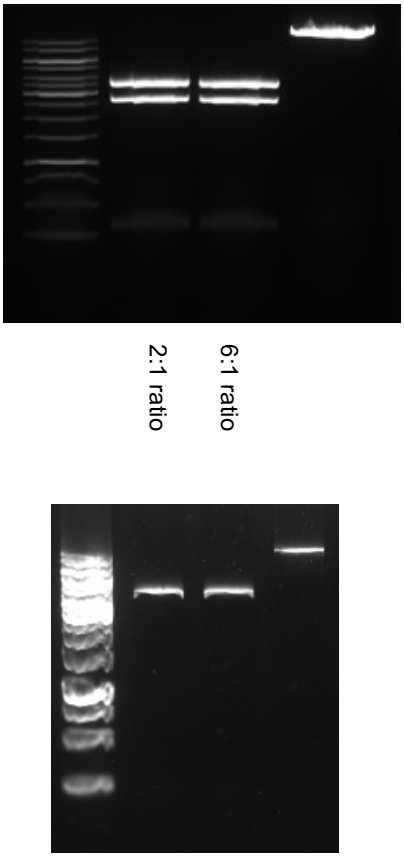
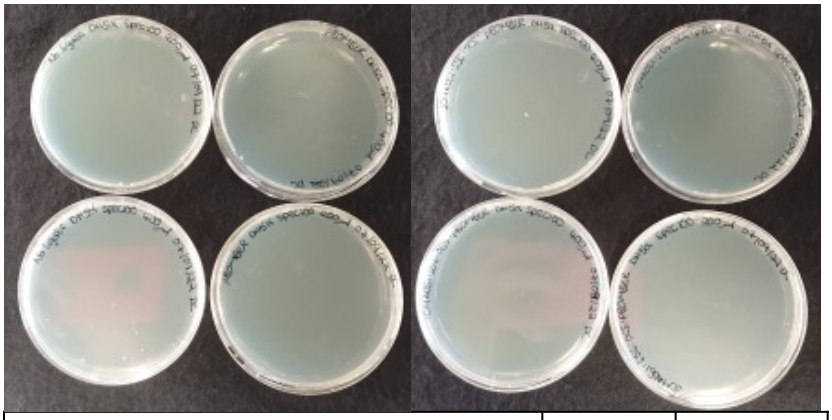


Figure 7: The above gel shows a digestion of *CMADS1* in pART7 mini-preps 1 and 3 with NotI with a pBOMBER digestion with NotI. The bottom gel shows the purified DNA of these digestions.

Insert:vector ratios were chosen based on the initial concentration calculation (2:1) and the purified DNA gel (6:1). For the ligation stage an overnight incubation at 20C was used for both ratios. 100 µl of competent cells was used for transformation and standard lab heat shock protocol followed. 200 µl and 400 µl of transformed cells in SOC media was spread of spectinomycin100 LBA plates for 24 hour 37C incubation.



	200 µl	400 µl
1 pBOMBER::*35S*::*CMADS1*	0	0
3 pBOMBER::*35S*::*CMADS1*	0	1
pBOMBER empty vector control	0	0
No ligase control	0	0

Figure 8: colony plates after 24 hour incubation showing only one colony growth in *35S::CMADS1* in pBOMBER.

The one colony was mini prepped using Isolate II mini prep kit and following the enclosed protocol. Double diagnostic digests were performed with no bands observed.

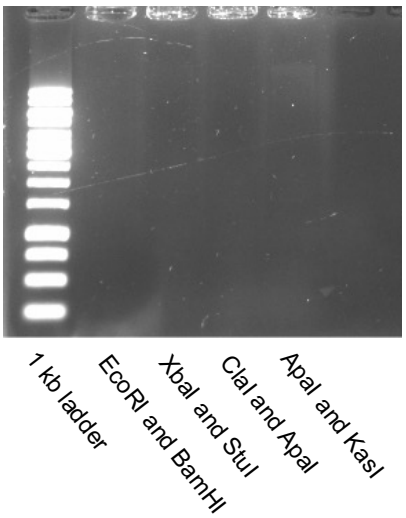
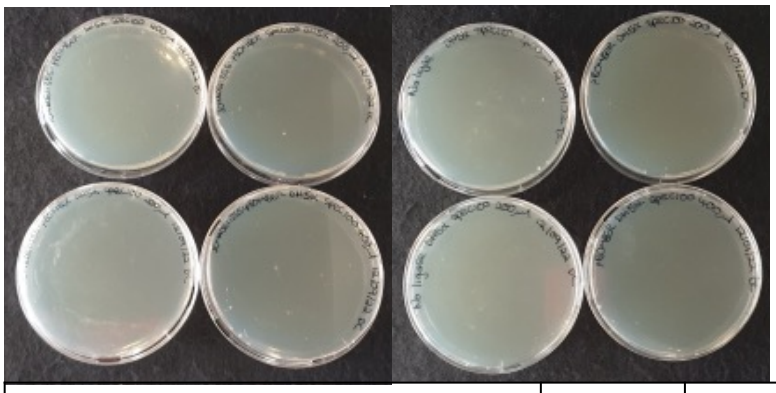


Figure 9: diagnostic digestion gel of *35S::CMADS1* in pBOMBER.

Transformations were re-done this time using a ligation incubation of 2 hours at 37C.



	200 µl	400 µl
1 CMADS1:35S:pBOMBER	0	0
3 CMADS1:35S:pBOMBER	1	4
pBOMBER empty vector control	0	0
No ligase control	0	0

Figure 10: colony plates of *35S::CMADS1* in pBOMBER with colony counts in the table.

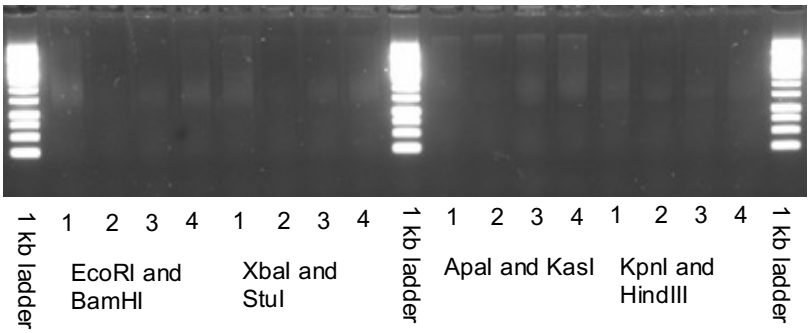


Figure 11: a diagnostic digest gel showing *35::CMADS1* in pBOMBER no bands were observed.

Agilent Ultracompetent XL10-Gold cells were purchased for their efficiency with larger (10 kb <) insertions, despite the insertions being approximately 3 kb. These provided

a higher colony growth but diagnostic digests showed unsuccessful insertions and the transformation was abandone